Compressive Sensing

Dr.-Ing. Ali Bereyhi

Friedrich-Alexander University

ali.bereyhi@fau.de

Summer 2022

Where Are We?

- *Optimal recovery is given by ℓ_0-norm minimization*
- *Best relaxation of optimal approach is ℓ_1-norm minimization*
  - *We learned various forms of ℓ_1-norm minimization*
  - *We also learned how to implement it*
- *We learned various iterative algorithms as well*
  - *Greedy algorithms: OMP, CoSaMP and Subspace Pursuit*
  - *Thresholding algorithms: IHT, HTP and IST*
- *We know how to handle sampling noise*
  - *We either extend the previous approaches via LS*
  - *We could alternatively follow Dantzig approach*

Where Are We?

But how exactly these algorithms work?

Of course, that depends on the sensing matrix!

- *For some matrix designs, one algorithm works fine*
- *For some others, it does not perform well*

But, we have not yet explored this degree of the freedom!

We discuss it in this part of the lecture

Sensing Matrices

Impacts of Good Sensing Matrix

The key point which distinguishes between sparse recovery and

Compressive Sensing

is the fact that in the latter, we are able to

design the sensing matrix

You might ask, but does it really matter?

Sure! Let's start with a simple example

Sensing Matrices

Section 1: An Initial Example

Impacts of Good Sensing Matrix: An Initial Example

Remember the OMP algorithm in noise-free case

- *Start with* $\boldsymbol{x}^{(0)} = \mathbf{0}$ *and* $\mathcal{S}^{(0)} = \emptyset$
- *Update the support for* $t \geq 1$ *as*

$$\mathcal{S}^{(t)} = \mathcal{S}^{(t-1)} \cup \operatorname*{argmax}_{n} \left| \left[\mathbf{A}^{\mathsf{T}} \left(\boldsymbol{y} - \mathbf{A}\boldsymbol{x}^{(t-1)} \right) \right]_{n} \right|$$

- *Set* $\boldsymbol{z} = \mathbf{A}^{\dagger}_{\mathcal{S}^{(t)}} \boldsymbol{y}$ *and update* $\boldsymbol{x}^{(t)}$ *as*

$$\boldsymbol{x}^{(t)}_{\mathcal{S}^{(t)}} = \boldsymbol{z}$$

- *Stop at iteration* T, *at which* $\mathbf{A}\boldsymbol{x}^{(T)} = \boldsymbol{y}$

Impacts of Good Sensing Matrix: An Initial Example

Also remember the key feature of OMP algorithm

> *The orthogonality principle indicates that*
>
> *In each iteration, a new index is added to the support*

Let us know see, if we could find some conditions on **A** *which*

lead to a recovery guarantee via OMP

To start, assume that

> *The original sparse signal is* $\boldsymbol{x}$ *with support* $\mathcal{S} = \operatorname{Supp}(\boldsymbol{x})$

Impacts of Good Sensing Matrix: An Initial Example

First of all, we need to make sure that the solution is unique

If we are given by the support $\mathcal{S}$,

We should be able to recover $\boldsymbol{x}$ uniquely

Knowing the support $\mathcal{S}$, the sparse recovery problem is

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{y}$$

$$\mathbf{A}_{\mathcal{S}}\boldsymbol{x}_{\mathcal{S}} = \boldsymbol{y}$$

Thus, the first condition is

$\mathbf{A}_{\mathcal{S}}$ *must be injective*

Impacts of Good Sensing Matrix: An Initial Example

In iteration 1, OMP should add an index from $\mathcal{S}$

$$\underset{n}{\operatorname{argmax}} \left| \left[\mathbf{A}^{\mathsf{T}} \left(\boldsymbol{y} - \mathbf{A}\boldsymbol{x}^{(0)} \right) \right]_n \right| \in \mathcal{S}$$

Since $\boldsymbol{x}^{(0)} = \mathbf{0}$, this means

$$\underset{n}{\operatorname{argmax}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{y} \right]_n \right| \in \mathcal{S}$$

This equivalently means we should have

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{y} \right]_n \right| > \max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{y} \right]_n \right|$$

Impacts of Good Sensing Matrix: An Initial Example

This should happen in each iteration

Assume, we are in iteration t and we have recovered up to now

$$\boldsymbol{x}^{(t-1)} : \operatorname{Supp}\left(\boldsymbol{x}^{(t-1)}\right) \subset \mathcal{S}$$

The residual would then read

$$\boldsymbol{r}^{(t-1)} = \boldsymbol{y} - \mathbf{A}\boldsymbol{x}^{(t-1)} = \mathbf{A}\boldsymbol{x} - \mathbf{A}\boldsymbol{x}^{(t-1)} = \mathbf{A}\left(\boldsymbol{x} - \boldsymbol{x}^{(t-1)}\right)$$

Since $\boldsymbol{x}$ and $\boldsymbol{x}^{(t-1)}$ are both supported on $\mathcal{S}$, we could say

$$\boldsymbol{r}^{(t-1)} = \mathbf{A}\boldsymbol{z} \text{ for some } \boldsymbol{z} \text{ whose support reads } \operatorname{Supp}(\boldsymbol{z}) \subseteq \mathcal{S}$$

Impacts of Good Sensing Matrix: An Initial Example

This should happen in each iteration

To make sure we choose an index in $\mathcal{S}$, we take the upper hand

$$\max_{n\in\mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right| > \max_{n\notin\mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right|$$

for any $\boldsymbol{r}$ which is in the set

$$\mathcal{U} = \{ \boldsymbol{r} = \mathbf{A}\boldsymbol{z} : \mathrm{Supp}\,(\boldsymbol{z}) \subseteq \mathcal{S} \}$$

Since $\boldsymbol{r}^{(t-1)} \in \mathcal{U}$, this constraint guarantees that

$$\max_{n\in\mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r}^{(t-1)} \right]_n \right| > \max_{n\notin\mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r}^{(t-1)} \right]_n \right|$$

Impacts of Good Sensing Matrix: An Initial Example

This should happen in each iteration

Or alternatively, this constraint guarantees that

$$\operatorname*{argmax}_{n\in\mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r}^{(t-1)} \right]_n \right| \in \mathcal{S}$$

Therefore, the second condition is

$$\max_{n\in\mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right| > \max_{n\notin\mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right|$$

for any $\boldsymbol{r} \in \mathcal{U}$

Impacts of Good Sensing Matrix: An Initial Example

So what we could conclude?

Recovery Guarantee for OMP

If for every $\mathcal{S} \subset \{1, \dots, N\}$ *with* $|\mathcal{S}| = s$, *we have*

1. $\mathbf{A}_\mathcal{S}$ *is injective*
2. *For all* $\boldsymbol{r} \in \mathcal{U}$, $\mathbf{A}$ *reads*

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right| > \max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right|$$

Then, OMP recovers any s-sparse signal sampled by $\mathbf{A}$

Well, it seems to be a bit tough! But yet explains impacts of $\mathbf{A}$

Sensing Matrices

Section 2: Characterizing Sensing Matrices

Characterizing Sensing Matrices

In general, for every algorithm, one can end with

multiple recovery guarantees

However, they are not necessarily easy to check!

A recovery guarantee usually looks like

If matrix $\mathbf{A}$ *satisfies*

Constraint $\mathcal{C}$ *with probability* P_0

Then, it recovers a s-sparse vector with probability P

Characterizing Sensing Matrices

Constraints of most well-known guarantees are given in terms of

- *Null space property*
- *Coherence of matrices*
- *Restricted isometry property*

We hence need to learn these definitions first!

We start with null space property

Null Space Property

What Is Null Space Property?

Consider sensing matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ *used for compressive sensing*

$\mathbf{A}$ *is a fat matrix*

Thus, we could say $\ker \mathbf{A} \neq \{\mathbf{0}\}$

$\boxed{\text{entries of } \ker \mathbf{A} \text{ are } N\text{-dimensional vectors}}$

We are going to use this a lot!

What Is Null Space Property?

Let us define few notations for the rest of our way

- $\bar{\mathcal{S}}$ *is the complement of* $\mathcal{S}$*, i.e.,*

  all entries that are in $\{1, \dots, N\}$ *but not in* $\mathcal{S}$

  Example: Say $N = 5$ *and* $\mathcal{S} = \{1, 4\}$*; then,*

$$\bar{\mathcal{S}} = \{2, 3, 5\}$$

What Is Null Space Property?

Let us now define few notations for the rest of our way

- *The complete $\mathcal{S}$-subvector of $\boldsymbol{v}$ is*

$$\tilde{\boldsymbol{v}}_{\mathcal{S}} = \begin{cases} \boldsymbol{v}_{\mathcal{S}} & \forall n \in \mathcal{S} \\ 0 & \forall n \in \bar{\mathcal{S}} \end{cases}$$

Example: Say $\mathcal{S} = \{1, 3\}$ and

$$\boldsymbol{v} = \begin{bmatrix} 2 \\ 4 \\ 3 \\ 2 \end{bmatrix} \rightsquigarrow \boldsymbol{v}_{\mathcal{S}} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \rightsquigarrow \tilde{\boldsymbol{v}}_{\mathcal{S}} = \begin{bmatrix} 2 \\ 0 \\ 3 \\ 0 \end{bmatrix}$$

What Is Null Space Property?

Clearly, we have the following properties

- *For any* $\boldsymbol{v}$, *we have*

$$\boldsymbol{v} = \tilde{\boldsymbol{v}}_{\mathcal{S}} + \tilde{\boldsymbol{v}}_{\bar{\mathcal{S}}}$$

Remember the example

$$\boldsymbol{v} = \begin{bmatrix} 2 \\ 4 \\ 3 \\ 2 \end{bmatrix} = \left(\tilde{\boldsymbol{v}}_{\mathcal{S}} = \begin{bmatrix} 2 \\ 0 \\ 3 \\ 0 \end{bmatrix} \right) + \left(\tilde{\boldsymbol{v}}_{\mathcal{S}} = \begin{bmatrix} 0 \\ 4 \\ 0 \\ 2 \end{bmatrix} \right)$$

What Is Null Space Property?

Clearly, we have the following properties

- *For any* $\boldsymbol{v}$ *and set* $\mathcal{S}$, *we have*

$$\|\tilde{\boldsymbol{v}}_{\mathcal{S}}\|_p = \|\boldsymbol{v}_{\mathcal{S}}\|_p$$

- *For vector* $\boldsymbol{v}$ *with support* $\mathcal{S} = \mathrm{Supp}\,(\boldsymbol{v})$, *we have*

$$\tilde{\boldsymbol{v}}_{\mathcal{S}} = \boldsymbol{v}$$

What Is Null Space Property?

Null Space Property on Support $\mathcal{S}$

On Support $\mathcal{S} \subset \{1, \dots, N\}$, $\mathbf{A}$ fulfills the null space property, if

$$\|\boldsymbol{v}_{\mathcal{S}}\|_1 < \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1$$

for all $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$

What does it mean?

The entries in $\mathcal{S}$ are not dominant in the null space

What would be the ideal case?

The ideal case would be $\|\boldsymbol{v}_{\mathcal{S}}\|_1 = 0$

What Is Null Space Property?

Null Space Property on Support $\mathcal{S}$

On Support $\mathcal{S} \subset \{1, \dots, N\}$, $\mathbf{A}$ fulfills the null space property, if

$$\|\boldsymbol{v}_{\mathcal{S}}\|_1 < \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1$$

for all $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$

But, why is it defined in terms of ℓ_1-norm?

Well! Because we want to discuss the basis pursuit algorithm

Shall we define it for other norms?

Sure! We could define it for any ℓ_p-norm

What Is Null Space Property?

ℓ_p Null Space Property on Support $\mathcal{S}$

On Support $\mathcal{S} \subset \{1, \dots, N\}$, $\mathbf{A}$ fulfills the ℓ_p null space property, if

$$\|\boldsymbol{v}_{\mathcal{S}}\|_p < \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_p$$

for all $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$

Pay attention to this point that ...

If we set $p = 0$; then, the property requires for all $\boldsymbol{v} \in \ker \mathbf{A}$

$$\boldsymbol{v}_{\mathcal{S}} \text{ be sparser than } \boldsymbol{v}_{\bar{\mathcal{S}}}$$

What Is Null Space Property?

ℓ_p Null Space Property

A *fulfills the ℓ_p null space property of order s, if*

It satisfies ℓ_p null space property on all supports of size s

This means, if we choose any $\mathcal{S}$ with s indices in it, we have

$$\|\boldsymbol{v}_{\mathcal{S}}\|_p < \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_p$$

for all $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$

Null Space Property

Section 1: Recovery Guarantee via Null Space Property

Why Null Space Property?

What is specific about the null space property?

> *We start with basis pursuit recovery from noise-free samples*

To this end, assume

> *We have s-sparse signal* $\boldsymbol{x} \in \mathbb{R}^N$ *with* $\mathrm{Supp}\left(\boldsymbol{x}\right) = \mathcal{S}$

We use basis pursuit to recover $\boldsymbol{x}$ *from its noise-free samples, i.e.,*

Basis Pursuit

$$\min_{\boldsymbol{z}} \|\boldsymbol{z}\|_1 \qquad \text{subject to } \mathbf{A}\boldsymbol{z} = \boldsymbol{y} = \mathbf{A}\boldsymbol{x}$$

Why Null Space Property?

What is specific about the null space property?

> *In general for any* $\boldsymbol{v} \in \ker \mathbf{A}$, *we have*
> $$\mathbf{A}\boldsymbol{v} = \mathbf{0}$$

So, if we set $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{v}$, *we have*

$$\mathbf{A}\left(\boldsymbol{x} + \boldsymbol{v}\right) = \boldsymbol{y}$$

So, we could say

> *all solutions to* $\mathbf{A}\boldsymbol{z} = \boldsymbol{y}$ *are of the form* $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{v}$

Why Null Space Property?

What is specific about the null space property?

In general for any $\boldsymbol{v} \in \ker \mathbf{A}$, *we have*

$$\mathbf{A}\boldsymbol{v} = \mathbf{0}$$

Now consider a solution $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{v}$ *for some* $\boldsymbol{v} \neq \mathbf{0}$

Before we start, note that $\boldsymbol{x} = \tilde{\mathbf{x}}_{\mathcal{S}}$ *and* $\tilde{\mathbf{x}}_{\bar{\mathcal{S}}} = \mathbf{0}$; *thus*

$$\tilde{\boldsymbol{z}}_{\bar{\mathcal{S}}} = \tilde{\boldsymbol{v}}_{\bar{\mathcal{S}}}$$

Keep this in mind for later!

Why Null Space Property?

What is specific about the null space property?

In general for any $\boldsymbol{v} \in \ker \mathbf{A}$, *we have*

$$\mathbf{A}\boldsymbol{v} = \mathbf{0}$$

Now consider a solution $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{v}$ *for some* $\boldsymbol{v} \neq \mathbf{0}$

Let's start with an obvious step

$$\|\boldsymbol{x}\|_1 = \|\boldsymbol{x} - \tilde{\boldsymbol{z}}_{\mathcal{S}} + \tilde{\boldsymbol{z}}_{\mathcal{S}}\|_1$$

Why Null Space Property?

What is specific about the null space property?

In general for any $\boldsymbol{v} \in \ker \mathbf{A}$, *we have*

$$\mathbf{A}\boldsymbol{v} = \mathbf{0}$$

Now consider a solution $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{v}$ *for some* $\boldsymbol{v} \neq \mathbf{0}$

We now use the triangle inequality and write

$$\begin{aligned}\|\boldsymbol{x}\|_1 &\leq \|\boldsymbol{x} - \tilde{\boldsymbol{z}}_{\mathcal{S}}\|_1 + \|\tilde{\boldsymbol{z}}_{\mathcal{S}}\|_1 \\ &= \|\tilde{\boldsymbol{v}}_{\mathcal{S}}\|_1 + \|\tilde{\boldsymbol{z}}_{\mathcal{S}}\|_1\end{aligned}$$

Why Null Space Property?

What is specific about the null space property?

In general for any $\boldsymbol{v} \in \ker \mathbf{A}$, *we have*

$$\mathbf{A}\boldsymbol{v} = \mathbf{0}$$

Now consider a solution $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{v}$ *for some* $\boldsymbol{v} \neq \mathbf{0}$

Null space property of $\mathbf{A}$ *on* $\mathcal{S}$ *indicate that* $\|\tilde{\boldsymbol{v}}_{\mathcal{S}}\|_1 < \|\tilde{\boldsymbol{v}}_{\bar{\mathcal{S}}}\|_1$

$$\begin{aligned} \|\boldsymbol{x}\|_1 &\leq \|\tilde{\boldsymbol{v}}_{\mathcal{S}}\|_1 + \|\tilde{\boldsymbol{z}}_{\mathcal{S}}\|_1 \\ &< \|\tilde{\boldsymbol{v}}_{\bar{\mathcal{S}}}\|_1 + \|\tilde{\boldsymbol{z}}_{\mathcal{S}}\|_1 \end{aligned}$$

Why Null Space Property?

What is specific about the null space property?

In general for any $\boldsymbol{v} \in \ker \mathbf{A}$*, we have*

$$\mathbf{A}\boldsymbol{v} = \mathbf{0}$$

Now consider a solution $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{v}$ *for some* $\boldsymbol{v} \neq \mathbf{0}$

But we know that $\tilde{\boldsymbol{z}}_{\bar{\mathcal{S}}} = \tilde{\boldsymbol{v}}_{\bar{\mathcal{S}}}$*; so, we could write*

$$\begin{aligned} \|\boldsymbol{x}\|_1 &< \|\tilde{\boldsymbol{z}}_{\bar{\mathcal{S}}}\|_1 + \|\tilde{\boldsymbol{z}}_{\mathcal{S}}\|_1 \\ &= \|\boldsymbol{z}\|_1 \end{aligned}$$

Why Null Space Property?

So, what have we concluded?

If $\mathbf{A}$ *fulfills the null space property on support of* $\boldsymbol{x}$,
$\boldsymbol{x}$ *is the unique solution to basis pursuit recovery*

So, we could equivalently say that

If $\mathbf{A}$ *fulfills the null space property of order s*
any s-sparse vector sampled by $\mathbf{A}$
is uniquely recovered via basis pursuit from its samples

Why Null Space Property?

If $\mathbf{A}$ *fulfills the null space property of order* s

any s*-sparse vector sampled by* $\mathbf{A}$

is uniquely recovered via basis pursuit from its samples

This proves the sufficiency of null space property, but . . .

Is it also necessary to have null-space property?

Why Null Space Property?

Let's check the necessity together

Assume for any $\boldsymbol{x}$ *with* $\operatorname{Supp}(\boldsymbol{x}) = \mathcal{S}$, *we know that*

$$\boldsymbol{x} = \operatorname{argmin} \|\boldsymbol{z}\|_1 \qquad \textit{subject to} \ \ \mathbf{A}\boldsymbol{z} = \mathbf{A}\boldsymbol{x}$$

and $\boldsymbol{x}$ *is unique*

This means if we take a vector $\boldsymbol{v}$, *and set* $\boldsymbol{x} = \tilde{\boldsymbol{v}}_{\mathcal{S}}$; *then,*

$$\tilde{\boldsymbol{v}}_{\mathcal{S}} = \operatorname{argmin} \|\boldsymbol{z}\|_1 \qquad \textit{subject to} \ \ \mathbf{A}\boldsymbol{z} = \mathbf{A}\tilde{\boldsymbol{v}}_{\mathcal{S}}$$

Why Null Space Property?

Let's check the necessity together

Assume for any $\mathbf{x}$ *with* $\operatorname{Supp}(\mathbf{x}) = \mathcal{S}$, *we know that*

$$\mathbf{x} = \operatorname{argmin} \|\mathbf{z}\|_1 \qquad \text{subject to } \mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}$$

and $\mathbf{x}$ *is unique*

Or equivalently, we could write

$$\|\tilde{\mathbf{v}}_{\mathcal{S}}\|_1 < \|\mathbf{z}\|_1$$

for all solutions of $\mathbf{A}\mathbf{z} = \mathbf{A}\tilde{\mathbf{v}}_{\mathcal{S}}$

Why Null Space Property?

We write the conclusion up to here below

If basis pursuit recovers $\boldsymbol{x}$ *sampled by* $\mathbf{A}$ *uniquely; then*

$$\|\tilde{\boldsymbol{v}}_{\mathcal{S}}\|_1 < \|\boldsymbol{z}\|_1$$

for all solutions of $\mathbf{A}\boldsymbol{z} = \mathbf{A}\tilde{\boldsymbol{v}}_{\mathcal{S}}$

Let us now set $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$*; then, we know*

$$\mathbf{0} = \mathbf{A}\boldsymbol{v} = \mathbf{A}\left(\tilde{\boldsymbol{v}}_{\mathcal{S}} + \tilde{\boldsymbol{v}}_{\bar{\mathcal{S}}}\right)$$

Or in other words, $\mathbf{A}\tilde{\boldsymbol{v}}_{\mathcal{S}} = -\mathbf{A}\tilde{\boldsymbol{v}}_{\bar{\mathcal{S}}}$*:* $\boldsymbol{z} = -\tilde{\boldsymbol{v}}_{\bar{\mathcal{S}}}$ *is a solution!*

Why Null Space Property?

This means that we could modify our conclusion as

If basis pursuit recovers $\boldsymbol{x}$ *sampled by* $\mathbf{A}$ *uniquely; then*

$$\|\tilde{\boldsymbol{v}}_{\mathcal{S}}\|_1 < \|\tilde{\boldsymbol{v}}_{\bar{\mathcal{S}}}\|_1$$

Or equivalently, one could say

If basis pursuit recovers $\boldsymbol{x}$ *sampled by* $\mathbf{A}$ *uniquely; then*

$\mathbf{A}$ *satisfies null space property on* $\mathcal{S} = \mathrm{Supp}\,(\boldsymbol{x})$

Null space property is a necessary condition too!

Why Null Space Property?

Recovery Guarantee by Null Space Property

Sparse signal $\boldsymbol{x}$ *is uniquely recovered via basis pursuit algorithm from its samples collected by sampling matrix* $\mathbf{A}$, *if and only if* $\mathbf{A}$ *satisfies null space property on* $\mathcal{S} = \operatorname{Supp}(\boldsymbol{x})$

This result also extends to ℓ_p*-norm minimization with* $0 < p \leq 1$

Recovery Guarantee by Null Space Property

Every s*-sparse signal* $\boldsymbol{x}$ *is uniquely recovered via* ℓ_p*-norm minimization from its samples collected by sampling matrix* $\mathbf{A}$, *if and only if* $\mathbf{A}$ *satisfies* ℓ_p *null space property of order* s

Null Space Property

Section 2: Robust Null Space Property

Robust Null Space Property

How could we extend the results to noisy sampling?

We need first to extend the definition of null space property

Robust Null Space Property on Support $\mathcal{S}$

On Support $\mathcal{S} \subset \{1, \dots, N\}$, $\mathbf{A}$ *fulfills robust null space property with respect to* ℓ_p*-norm and with constants* ρ *and* τ*, if*

$$\|\boldsymbol{v}_{\mathcal{S}}\|_1 \leq \rho \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1 + \tau \|\mathbf{A}\boldsymbol{v}\|_p$$

for all $\boldsymbol{v} \in \mathbb{R}^N$, $0 < \rho \leq 1$ *and* $\tau \geq 0$

Robust Null Space Property

Similarly, we can define robust null space property

Robust Null Space Property

A *fulfills robust null space property of order s with respect to ℓ_p-norm and with constants ρ and τ, if it satisfies robust null space property on all supports of size s*

Recovery Guarantee for Noisy Sparse Recovery

We now extend the results to noisy sampling

$$\text{We have } s\text{-sparse signal } \mathbf{x} \in \mathbb{R}^N \text{ with } \mathrm{Supp}(\mathbf{x}) = \mathcal{S}$$

We collect noisy samples via sampling matrix $\mathbf{A}$, *i.e.,*

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$$

We use basis pursuit with quadratic constraint for sparse recovery

Basis Pursuit with Quadratic Constraint

$$\min_{\mathbf{z}} \|\mathbf{z}\|_1 \qquad \text{subject to} \quad \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \eta$$

Recovery Guarantee for Noisy Sparse Recovery

Let's start with a simply identity

$$\begin{aligned} \|\boldsymbol{x}\|_1 &= \|\tilde{\boldsymbol{x}}_{\mathcal{S}}\|_1 + \|\tilde{\boldsymbol{x}}_{\bar{\mathcal{S}}}\|_1 \\ &= \|\tilde{\boldsymbol{x}}_{\mathcal{S}}\|_1 + 0 \\ &= \|\boldsymbol{x}_{\mathcal{S}}\|_1 \end{aligned}$$

For an arbitrary signal $\boldsymbol{z}$*, we can always write*

$$\begin{aligned} \|\boldsymbol{x}_{\mathcal{S}}\|_1 &= \|(\boldsymbol{x} - \boldsymbol{z} + \boldsymbol{z})_{\mathcal{S}}\|_1 \\ &= \|\boldsymbol{x}_{\mathcal{S}} - \boldsymbol{z}_{\mathcal{S}} + \boldsymbol{z}_{\mathcal{S}}\|_1 \\ &\leq \|\boldsymbol{x}_{\mathcal{S}} - \boldsymbol{z}_{\mathcal{S}}\|_1 + \|\boldsymbol{z}_{\mathcal{S}}\|_1 \end{aligned}$$

Recovery Guarantee for Noisy Sparse Recovery

Once again use the triangle inequality and write

$$\begin{aligned}\|(\boldsymbol{x}-\boldsymbol{z})_{\bar{\mathcal{S}}}\|_1 &\leq \|\boldsymbol{x}_{\bar{\mathcal{S}}}\|_1 + \|\boldsymbol{z}_{\bar{\mathcal{S}}}\|_1 \\ &= \|\boldsymbol{z}_{\bar{\mathcal{S}}}\|_1\end{aligned}$$

So we have concluded that

$$\begin{aligned}\|(\boldsymbol{x}-\boldsymbol{z})_{\bar{\mathcal{S}}}\|_1 &\leq \|\boldsymbol{z}_{\bar{\mathcal{S}}}\|_1 \\ \|\boldsymbol{x}\|_1 &\leq \|(\boldsymbol{x}-\boldsymbol{z})_{\mathcal{S}}\|_1 + \|\boldsymbol{z}_{\mathcal{S}}\|_1\end{aligned}$$

By adding up these two inequalities, we have

$$\|(\boldsymbol{x}-\boldsymbol{z})_{\bar{\mathcal{S}}}\|_1 + \|\boldsymbol{x}\|_1 \leq \|\boldsymbol{z}_{\bar{\mathcal{S}}}\|_1 + \|(\boldsymbol{x}-\boldsymbol{z})_{\mathcal{S}}\|_1 + \|\boldsymbol{z}_{\mathcal{S}}\|_1$$

Recovery Guarantee for Noisy Sparse Recovery

By sorting up the inequality, we conclude that

$$\|(\boldsymbol{x} - \boldsymbol{z})_{\bar{\mathcal{S}}}\|_1 \leq \|\boldsymbol{z}_{\bar{\mathcal{S}}}\|_1 + \|(\boldsymbol{x} - \boldsymbol{z})_{\mathcal{S}}\|_1 + \|\boldsymbol{z}_{\mathcal{S}}\|_1 - \|\boldsymbol{x}\|_1$$

Let us call $\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{z}$*; so we have*

$$\begin{aligned}\|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1 &\leq \|\boldsymbol{v}_{\mathcal{S}}\|_1 + \underbrace{\|\boldsymbol{z}_{\bar{\mathcal{S}}}\|_1 + \|\boldsymbol{z}_{\mathcal{S}}\|_1}_{\|\boldsymbol{z}\|_1} - \|\boldsymbol{x}\|_1 \\ &\leq \|\boldsymbol{v}_{\mathcal{S}}\|_1 + \|\boldsymbol{z}\|_1 - \|\boldsymbol{x}\|_1\end{aligned}$$

W are going to use this inequality later on!

Recovery Guarantee for Noisy Sparse Recovery

Now assume that $\mathbf{A}$ *fulfills*

$$\boxed{\textit{robust null space property on } \mathcal{S}}$$

We want to find out

How much an estimation $\boldsymbol{z}$ *is distanced from the true signal* $\boldsymbol{x}$*?*

Let us define the error signal $\boldsymbol{v}$ *as*

$$\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{z}$$

Recovery Guarantee for Noisy Sparse Recovery

$\boxed{\mathbf{A} \textit{ fulfills robust null space property on } \mathcal{S}}$

From the definition, we could write

$$\|\boldsymbol{v}_{\mathcal{S}}\|_1 \leq \rho \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1 + \tau \|\mathbf{A}\boldsymbol{v}\|_p$$

Moreover, we just showed that

$$\|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1 \leq \|\boldsymbol{v}_{\mathcal{S}}\|_1 + \|\boldsymbol{z}\|_1 - \|\boldsymbol{x}\|_1$$

Recovery Guarantee for Noisy Sparse Recovery

Combining these two inequalities, we have

$$\|\boldsymbol{v}_{\mathcal{S}}\|_1 + \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1 \leq \rho \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1 + \tau \|\mathbf{A}\boldsymbol{v}\|_p + \|\boldsymbol{v}_{\mathcal{S}}\|_1 + \|\boldsymbol{z}\|_1 - \|\boldsymbol{x}\|_1$$
$$\|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1 \leq \rho \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1 + \|\boldsymbol{z}\|_1 - \|\boldsymbol{x}\|_1 + \tau \|\mathbf{A}\boldsymbol{v}\|_p$$

So we could write

$$(1-\rho)\,\|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1 \leq \|\boldsymbol{z}\|_1 - \|\boldsymbol{x}\|_1 + \tau \|\mathbf{A}\boldsymbol{v}\|_p$$

Or equivalently, we have

$$\|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1 \leq \frac{1}{1-\rho}\left(\|\boldsymbol{z}\|_1 - \|\boldsymbol{x}\|_1 + \tau \|\mathbf{A}\boldsymbol{v}\|_p\right)$$

Recovery Guarantee for Noisy Sparse Recovery

We could once again, combine

$$\|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1 \leq \frac{1}{1-\rho}\left(\|\boldsymbol{z}\|_1 - \|\boldsymbol{x}\|_1 + \tau\|\mathbf{A}\boldsymbol{v}\|_p\right)$$

with the definition of robust null space property, i.e.,

$$\|\boldsymbol{v}_{\mathcal{S}}\|_1 \leq \rho\|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1 + \tau\|\mathbf{A}\boldsymbol{v}\|_p$$

Recovery Guarantee for Noisy Sparse Recovery

Combining these two inequalities leads to

$$\begin{aligned}\|\mathbf{v}\|_1 &= \|\mathbf{v}_{\bar{\mathcal{S}}}\|_1 + \|\mathbf{v}_{\mathcal{S}}\|_1 \\ &\leq \|\mathbf{v}_{\bar{\mathcal{S}}}\|_1 + \rho\|\mathbf{v}_{\bar{\mathcal{S}}}\|_1 + \tau\|\mathbf{A}\mathbf{v}\|_p \\ &= (1+\rho)\,\|\mathbf{v}_{\bar{\mathcal{S}}}\|_1 + \tau\|\mathbf{A}\mathbf{v}\|_p \\ &\leq \frac{1+\rho}{1-\rho}\left(\|\mathbf{z}\|_1 - \|\mathbf{x}\|_1 + \tau\|\mathbf{A}\mathbf{v}\|_p\right) + \tau\|\mathbf{A}\mathbf{v}\|_p\end{aligned}$$

Recovery Guarantee for Noisy Sparse Recovery

Since we know that

$$\frac{1+\rho}{1-\rho} \geq 1$$

We could finally write

$$\|\boldsymbol{v}\|_1 \leq \frac{1+\rho}{1-\rho}\left(\|\boldsymbol{z}\|_1 - \|\boldsymbol{x}\|_1 + 2\tau\|\mathbf{A}\boldsymbol{v}\|_p\right)$$

Recovery Guarantee for Noisy Sparse Recovery

What do we conclude now?

If $\mathbf{A}$ *fulfills robust null space property on* $\mathcal{S}$*; then,*

The recovery error is bounded

Just assume, we find some $\boldsymbol{z}$ *close to* $\boldsymbol{x}$*; then,*

$$\|\boldsymbol{z}\|_1 \approx \|\boldsymbol{x}\|_1$$

and thus

$$\|\boldsymbol{v}\|_1 \leq \epsilon$$

Recovery Guarantee for Noisy Sparse Recovery

The robust null space property, is a sufficient condition

for the given inequality

But, can we also show its necessity?

Yes! You get it as homework

Recovery Guarantee with Noisy Sampling: Conclusion One

Necessity and Sufficiency of Robust Null Space Property

$\mathbf{A}$ *satisfies robust null space property on* $\mathcal{S} = \mathrm{Supp}(\boldsymbol{x})$ *with parameters* $0 < \rho \leq 1$ *and* $\tau \geq 0$, *if and only if, for any* $\boldsymbol{z}$, *we have*

$$\|\boldsymbol{v}\|_1 \leq \frac{1+\rho}{1-\rho}\left(\|\boldsymbol{z}\|_1 - \|\boldsymbol{x}\|_1 + 2\tau\|\mathbf{A}\boldsymbol{v}\|_p\right)$$

where $\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{z}$

We now use this result to prove a recovery guarantee!

Recovery Guarantee with Noisy Sampling

Now let us assume that $\boldsymbol{x}^\sharp$ *is given by*

Basis Pursuit with Quadratic Constraint

$$\boldsymbol{x}^\sharp = \operatorname*{argmin}_{\boldsymbol{z}} \|\boldsymbol{z}\|_1 \qquad \textit{subject to} \;\; \|\mathbf{A}\boldsymbol{z} - \boldsymbol{y}\|_2 \leq \eta$$

Also assume that

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{x} + \boldsymbol{w}$$

where $\|\boldsymbol{w}\|_2 \leq \eta$

Recovery Guarantee with Noisy Sampling

Now let us assume that $\mathbf{x}^\sharp$ *is given by*

Basis Pursuit with Quadratic Constraint

$$\mathbf{x}^\sharp = \underset{\mathbf{z}}{\operatorname{argmin}} \|\mathbf{z}\|_1 \qquad \textit{subject to} \quad \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \eta$$

First of all, we note that

$$\begin{aligned} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 &= \|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{x} - \mathbf{w}\|_2 \\ &= \|\mathbf{w}\|_2 \leq \eta \end{aligned}$$

This means that $\mathbf{x}$ *is also a feasible point!*

Recovery Guarantee with Noisy Sampling

Now let us assume that $\mathbf{x}^\sharp$ *is given by*

Basis Pursuit with Quadratic Constraint

$$\mathbf{x}^\sharp = \operatorname*{argmin}_{\mathbf{z}} \|\mathbf{z}\|_1 \qquad \textit{subject to} \;\; \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \leq \eta$$

Since $\mathbf{x}^\sharp$ *is the minimizer, we have*

$$\|\mathbf{x}^\sharp\|_1 \leq \|\textit{any feasible } \mathbf{z}\|_1 \rightsquigarrow \|\mathbf{x}^\sharp\|_1 \leq \|\mathbf{x}\|_1$$

Recovery Guarantee with Noisy Sampling

Reminder: *Conclusion One*

$\mathbf{A}$ *satisfies robust null space property on* $\mathcal{S} = \operatorname{Supp}(\boldsymbol{x})$, *if and only if, for any* $\boldsymbol{z}$, *we have*

$$\|\boldsymbol{v}\|_1 \leq \frac{1+\rho}{1-\rho}\left(\|\boldsymbol{z}\|_1 - \|\boldsymbol{x}\|_1 + 2\tau\|\mathbf{A}\boldsymbol{v}\|_p\right)$$

where $\boldsymbol{v} = \boldsymbol{x} - \boldsymbol{z}$

We now use Conclusion One and set $\boldsymbol{z} = \boldsymbol{x}^\sharp$

$$\|\boldsymbol{x} - \boldsymbol{x}^\sharp\|_1 \leq \frac{1+\rho}{1-\rho}\left(\|\boldsymbol{x}^\sharp\|_1 - \|\boldsymbol{x}\|_1 + 2\tau\left\|\mathbf{A}\left(\boldsymbol{x} - \boldsymbol{x}^\sharp\right)\right\|_p\right)$$

Recovery Guarantee with Noisy Sampling

$$\boxed{\|\boldsymbol{x} - \boldsymbol{x}^\sharp\|_1 \leq \frac{1+\rho}{1-\rho} \left(\|\boldsymbol{x}^\sharp\|_1 - \|\boldsymbol{x}\|_1 + 2\tau \|\mathbf{A}\left(\boldsymbol{x} - \boldsymbol{x}^\sharp\right)\|_p \right)}$$

First let's use the fact that

$$\|\boldsymbol{v}\|_2 \leq \|\boldsymbol{v}\|_1$$

Try at home to show this!

So, we conclude that

$$\|\boldsymbol{x} - \boldsymbol{x}^\sharp\|_2 \leq \frac{1+\rho}{1-\rho} \left(\|\boldsymbol{x}^\sharp\|_1 - \|\boldsymbol{x}\|_1 + 2\tau \|\mathbf{A}\left(\boldsymbol{x} - \boldsymbol{x}^\sharp\right)\|_p \right)$$

Recovery Guarantee with Noisy Sampling

$$\boxed{\|\boldsymbol{x} - \boldsymbol{x}^\sharp\|_2 \leq \frac{1+\rho}{1-\rho}\left(\|\boldsymbol{x}^\sharp\|_1 - \|\boldsymbol{x}\|_1 + 2\tau \|\mathbf{A}\left(\boldsymbol{x} - \boldsymbol{x}^\sharp\right)\|_p \right)}$$

As $\|\boldsymbol{x}^\sharp\|_1 \leq \|\boldsymbol{x}\|_1$, we could conclude that

$$\|\boldsymbol{x}^\sharp\|_1 - \|\boldsymbol{x}\|_1 \leq 0$$

So, we conclude that

$$\|\boldsymbol{x} - \boldsymbol{x}^\sharp\|_2 \leq \frac{1+\rho}{1-\rho} 2\tau \|\mathbf{A}\left(\boldsymbol{x} - \boldsymbol{x}^\sharp\right)\|_p$$

Recovery Guarantee with Noisy Sampling

$$\boxed{\|\mathbf{x} - \mathbf{x}^\sharp\|_2 \leq 2\tau \frac{1+\rho}{1-\rho} \|\mathbf{A}\left(\mathbf{x} - \mathbf{x}^\sharp\right)\|_p}$$

Since $\mathbf{x}^\sharp$ is a feasible point, we have

$$\begin{aligned}
\eta \geq \|\mathbf{A}\mathbf{x}^\sharp - \mathbf{y}\|_2 &= \|\mathbf{A}\mathbf{x}^\sharp - \mathbf{A}\mathbf{x} - \mathbf{w}\|_2 \\
&= \|\mathbf{A}\left(\mathbf{x}^\sharp - \mathbf{x}\right) + \mathbf{w}\|_2 \\
&\geq \|\mathbf{A}\left(\mathbf{x}^\sharp - \mathbf{x}\right)\|_2 - \|\mathbf{w}\|_2 \\
&\geq \|\mathbf{A}\left(\mathbf{x}^\sharp - \mathbf{x}\right)\|_2 - \eta
\end{aligned}$$

Recovery Guarantee with Noisy Sampling

$$\boxed{\|\boldsymbol{x} - \boldsymbol{x}^{\sharp}\|_2 \leq 2\tau \frac{1+\rho}{1-\rho} \|\mathbf{A}\left(\boldsymbol{x} - \boldsymbol{x}^{\sharp}\right)\|_p}$$

So, we could write

$$\eta \geq \|\mathbf{A}\left(\boldsymbol{x}^{\sharp} - \boldsymbol{x}\right)\|_2 - \eta$$

Or equivalently

$$\|\mathbf{A}\left(\boldsymbol{x}^{\sharp} - \boldsymbol{x}\right)\|_2 \leq 2\eta$$

Recovery Guarantee with Noisy Sampling

$$\boxed{\|\mathbf{x} - \mathbf{x}^\sharp\|_2 \leq 2\tau \frac{1+\rho}{1-\rho} \|\mathbf{A}\left(\mathbf{x} - \mathbf{x}^\sharp\right)\|_p}$$

So, if we set $p = 2$, *we have*

$$\|\mathbf{x} - \mathbf{x}^\sharp\|_2 \leq \frac{4\tau\eta\left(1+\rho\right)}{1-\rho}$$

Recovery Guarantee with Noisy Sampling: Conclusion Two

Recovery Guarantee with Noisy Sampling

Let sparse signal $\boldsymbol{x}$ *be sampled by* $\mathbf{A}$ *which satisfies robust null space property on* $\mathcal{S} = \mathrm{Supp}\left(\boldsymbol{x}\right)$ *with parameters* $0 < \rho < 1$ *and* $\tau > 0$*. Assume noisy samples are given to basis pursuit with quadratic constraint* $\eta \geq \|\boldsymbol{w}\|_2$*. Then, recovered signal* $\boldsymbol{x}^\sharp$ *reads*

$$\|\boldsymbol{x} - \boldsymbol{x}^\sharp\|_2 \leq \frac{4\tau\eta\left(1+\rho\right)}{1-\rho}$$

Recovery Guarantee with Noisy Sampling: Conclusion Two

What happens if we send $\eta \to 0$?

We have

$$\|\boldsymbol{x} - \boldsymbol{x}^{\sharp}\|_2 \leq 0$$

In other words

$$\boldsymbol{x} = \boldsymbol{x}^{\sharp}$$

It guarantees exact recovery, i.e., it reduces to

recovery guarantee for basis pursuit with noise-free samples

Summary

What did we learn about null space property

If sampling matrix satisfies null space property

- *With noise-free samples*

  Basis pursuit successfully does exact recovery

- *With noisy samples*

  Basis pursuit finds a very good quality estimation

But can we really check the null space property?

Complexity of Checking Null Space Property

To check if the null space property of order s holds for $\mathbf{A}$

$$\boxed{\textit{We should check all subsets of size } s}$$

This means, we need to check

$$\binom{N}{s} \text{ cases}$$

which has the same complexity as the ℓ_0-norm minimization!

Coherence of Matrices

An Agreement

Before we start with definitions, let us make an agreement

From now on, we always assume that

sampling matrix has columns with unit ℓ_2-norm

unless we indicate the opposite

This means that

$$\mathbf{A} = [\boldsymbol{a}_1, \dots \boldsymbol{a}_N] \quad \text{with} \quad \|\boldsymbol{a}_n\|_2 = 1$$

What is Coherence?

Coherence

For matrix $\mathbf{A}$ *with* N *unit-norm column vectors, the coherence is the maximum inner product of two different columns*

$$\mu(\mathbf{A}) = \max_{1 \leq i \neq j \leq N} \left| \boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_j \right|$$

What is Coherence?

Attention: *We assumed that* $\mathbf{A}$ *has unit-norm columns!*

With non-normalized columns, we should further normalize . . .

Coherence of a Generic Matrix

For matrix $\mathbf{A}$ *with* N *column vectors, the coherence is the maximum inner product of two different columns*

$$\mu(\mathbf{A}) = \max_{1 \leq i \neq j \leq N} \frac{\left|\boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_j\right|}{\|\boldsymbol{a}_i\|_2 \|\boldsymbol{a}_j\|_2}$$

What is Coherence?

Coherence

For matrix $\mathbf{A}$ *with* N *unit-norm column vectors, the coherence is the maximum inner product of two different columns*

$$\mu(\mathbf{A}) = \max_{1 \leq i \neq j \leq N} \left| \boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_j \right|$$

Note that in compressive sensing $\mathbf{A}$ *is fat*

This means some of the columns are linearly dependent

What is Coherence?

Coherence

For matrix $\mathbf{A}$ *with* N *unit-norm column vectors, the coherence is the maximum inner product of two different columns*

$$\mu(\mathbf{A}) = \max_{1 \leq i \neq j \leq N} \left| \boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_j \right|$$

Coherence intuitively gives

the maximum correlation among all columns

ℓ_1 Coherence Function

We could moreover calculate the sum-correlation over a support

Sum-Correlation

Consider $\mathbf{A}$ *with* N *unit-norm column vectors, the sum-correlation of column* i *over support* $\mathcal{S}$ *is*

$$\sigma_i(\mathbf{A})_{\mathcal{S}} = \sum_{j \in \mathcal{S}} \left| \boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_j \right|$$

Sum-correlation calculates total correlation of columns in $\mathcal{S}$ *with* $\boldsymbol{a}_i$

ℓ_1 Coherence Function

Note that coherence $\mu(\mathbf{A})$ is maximum pair-wise correlation, i.e.,

$$\left| \boldsymbol{a}_i^\mathsf{T} \boldsymbol{a}_j \right| \leq \mu(\mathbf{A})$$

for any $i \neq j$; thus, we could say if $i \notin \mathcal{S}$, we have

$$\begin{aligned} \sigma_i(\mathbf{A})_{\mathcal{S}} &= \sum_{j \in \mathcal{S}} \left| \boldsymbol{a}_i^\mathsf{T} \boldsymbol{a}_j \right| \\ &\leq \sum_{j \in \mathcal{S}} \mu(\mathbf{A}) = |\mathcal{S}| \, \mu(\mathbf{A}) \end{aligned}$$

Keep it in mind!

ℓ_1 Coherence Function

Maximal Sum-Coherence

Consider **A** *with N column vectors, maximal sum-correlation for sparsity s is*

$$\sigma_i^{\max}(s) = \max\left\{\sigma_i(\mathbf{A})_{\mathcal{S}} : \mathcal{S} \subset \{1, \ldots, N\}, |\mathcal{S}| = s, i \notin \mathcal{S}\right\}$$

So, we check all subsets of size s and find the one

over which sum-correlation is maximal

ℓ_1 Coherence Function

Finally, we can give a more general definition of coherence

ℓ_1 Coherence Function

Consider $\mathbf{A}$ *with* N *columns,* ℓ_1 *coherence function for sparsity* s *is*

$$\mu_1(s) = \max_i \sigma_i^{\max}(s)$$

So ℓ_1 *coherence function chooses the column whose*

Maximal sum-correlation is is maximum

ℓ_1 Coherence Function

Finally, we can give a more general definition of coherence

ℓ_1 Coherence Function

Consider $\mathbf{A}$ *with* N *column,* ℓ_1 *coherence function for sparsity* s *is*

$$\mu_1(s) = \max_i \sigma_i^{\max}(s)$$

Why we call it ℓ_1 *coherence function? Since*

$$\sigma_i(\mathbf{A})_{\mathcal{S}} = \sum_{j \in \mathcal{S}} \left| \boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_j \right|$$

looks like the ℓ_1*-norm*

ℓ_1 Coherence Function

But ℓ_1 coherence function is still complex to calculate!

Well, Yes! But we can always bound it by coherence

How can we do this?

Remember that

$$\sigma_i\left(\mathbf{A}\right)_{\mathcal{S}} \leq |\mathcal{S}|\,\mu\left(\mathbf{A}\right)$$

So this is also true for maximal sum-correlation, i.e.,

$$\sigma_i^{\max}\left(s\right) \leq |\mathcal{S}|\,\mu\left(\mathbf{A}\right) = s\mu\left(\mathbf{A}\right)$$

ℓ_1 Coherence Function

But ℓ_1 coherence function is still complex to calculate!

> *Well, Yes! But we can always bound it by coherence*

How can we do this?

> *And also true when we maximize it over i, i.e.,*
>
> $$\mu_1(s) \leq s\mu(\mathbf{A})$$

ℓ_1 Coherence Function

But ℓ_1 coherence function is still complex to calculate!

Well, Yes! But we can always bound it by coherence

How can we do this?

Since all correlation terms are non-negative, we could say

$\mu_1(s)$ *definitely includes maximum correlation, i.e.,*

$$\mu_1(s) = \mu(\mathbf{A}) + \sum \dots$$
$$\geq \mu(\mathbf{A})$$

ℓ_1 Coherence Function

But ℓ_1 coherence function is still complex to calculate!

Well, Yes! But we can always bound it by coherence

How can we do this?

So, we can conclude that

$$\mu(\mathbf{A}) \leq \mu_1(s) \leq s\mu(\mathbf{A})$$

We are going to use these bounds a lot

Coherence of Matrices

Section 1: Good Values for Coherence

Coherence of Invertible Matrices

Bounding Spectrum of Matrices

Assume that ℓ_1 *coherence function of* $\mathbf{A}$ *is* $\mu_1(s)$*. Then, for any support* $\mathcal{S} \subset \{1, \dots, N\}$ *with* $|\mathcal{S}| = s$*, eigenvalues of* $\mathbf{A}_{\mathcal{S}}^{\mathsf{T}} \mathbf{A}_{\mathcal{S}}$ *read*

$$1 - \mu_1(s-1) \leq \lambda \leq 1 + \mu_1(s-1)$$

How can we prove this?

Via Gershgorin's disk theorem, Appendix A of the textbook

You will get an assignment for that!

Coherence of Invertible Matrices

What does this result say?

Let's consider the simple noise-free case

Assume we sample sparse signal $\mathbf{x}$ *with* $\mathcal{S} = \mathrm{Supp}\,(\mathbf{x})$

To guarantee unique recovery, we need to make sure that

$$\mathbf{A}_{\mathcal{S}}^{\mathsf{T}}\mathbf{A}_{\mathcal{S}} \text{ is invertible}$$

One might wonder why?

We must be able to invert $\mathbf{A}$ *at least when we know* $\mathcal{S}$

Coherence of Invertible Matrices

What does this result say?

$\boxed{\mathbf{A}_{\mathcal{S}}^{\mathsf{T}}\mathbf{A}_{\mathcal{S}} \textit{ should be invertible}}$

This means all eigenvalues of $\mathbf{A}_{\mathcal{S}}^{\mathsf{T}}\mathbf{A}_{\mathcal{S}}$ *should be positive, i.e.,*

$$1 - \mu_1 (s-1) > 0 \rightsquigarrow \mu_1 (s-1) < 1$$

Since $\mu_1 (s-1) \leq (s-1)\, \mu(\mathbf{A})$, *we are sure* $\mu_1 (s-1) < 1$, *if*

$$(s-1)\, \mu(\mathbf{A}) < 1 \rightsquigarrow \mu(\mathbf{A}) < \frac{1}{s-1}$$

Good Coherence: Observation One

Submatrix Invertibility

Any submatrix of size s of matrix $\mathbf{A}$ is injective, if

$$\mu(\mathbf{A}) < \frac{1}{s-1}$$

Well, what do we conclude

To be able to do sparse recovery, we should make sure that

$\mathbf{A}$ *has small coherence!*

Coherence of Matrices

Section 2: Matrices with Small Coherence

Small Coherence

How can we make the coherence small?

Intuitively, we should increase the number of rows

- *More rows means longer column vectors*
- *Larger dimension results in more orthogonal columns*
- *More orthogonality makes coherence smaller*

But can we show it rigorously?

Sure! We need review few definitions and properties first

Some Basics

Froebenuis Norm

For matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, the Froebenuis norm is defined as

$$\|\mathbf{A}\|_F = \sqrt{\sum_{n=1}^{N} \sum_{m=1}^{M} |a_{m,n}|^2}$$

where $a_{m,n}$ is the entry of $\mathbf{A}$ at row m and column n

You could show as a homework that

$$\|\mathbf{A}\|_F = \sqrt{\operatorname{tr}\left\{\mathbf{A}\mathbf{A}^\mathsf{T}\right\}}$$

Some Basics

Cauchy-Schwarz Inequality

For matrices $\mathbf{A} \in \mathbb{R}^{M \times N}$ *and* $\mathbf{B} \in \mathbb{R}^{N \times K}$, *we have*

$$\mathrm{tr}\left\{\mathbf{AB}\right\} \leq \|\mathbf{A}\|_F \|\mathbf{B}\|_F$$

Cauchy-Schwarz is actually a much general inequality

Check it in the Wikipedia page

Some Basics

Circular Shift Property of Trace

For matrices $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{D}$, *with matching dimensions, we have*

$$\operatorname{tr}\{\mathbf{ABD}\} = \operatorname{tr}\{\mathbf{DAB}\} = \operatorname{tr}\{\mathbf{BDA}\}$$

We can always shift circularly the multiplied arguments of trace

as long as the dimensions match

Small Coherence

How can we make the coherence small?

Let us start with a simple identity

$$\mathbf{A}\mathbf{A}^{\mathsf{T}} = \mathbf{A}\mathbf{A}^{\mathsf{T}}\mathbf{I}_M$$

where $\mathbf{I}_M$ *is* $M \times M$ *identity. Now, we use Cauchy-Schwarz*

$$\operatorname{tr}\left\{\mathbf{A}\mathbf{A}^{\mathsf{T}}\mathbf{I}_M\right\} \leq \|\mathbf{A}\mathbf{A}^{\mathsf{T}}\|_F \|\mathbf{I}_M\|_F$$

Thus, we have

$$\operatorname{tr}\left\{\mathbf{A}\mathbf{A}^{\mathsf{T}}\right\} \leq \|\mathbf{A}\mathbf{A}^{\mathsf{T}}\|_F \|\mathbf{I}_M\|_F$$

Small Coherence

How can we make the coherence small?

We know that $\|\mathbf{I}_M\|_F^2 = M$*; thus,*

$$\operatorname{tr}\left\{\mathbf{A}\mathbf{A}^\mathsf{T}\right\} \leq \sqrt{M}\|\mathbf{A}\mathbf{A}^\mathsf{T}\|_F$$

Now we calculate $\|\mathbf{A}\mathbf{A}^\mathsf{T}\|_F$

Start with the fact that

$$\|\mathbf{A}\mathbf{A}^\mathsf{T}\|_F^2 = \operatorname{tr}\left\{\mathbf{A}\mathbf{A}^\mathsf{T}\left(\mathbf{A}\mathbf{A}^\mathsf{T}\right)^\mathsf{T}\right\}$$

Small Coherence

How can we make the coherence small?

We know that $\|\mathbf{I}_M\|_F^2 = M$*; thus,*

$$\mathrm{tr}\left\{\mathbf{A}\mathbf{A}^\mathsf{T}\right\} \leq \sqrt{M}\|\mathbf{A}\mathbf{A}^\mathsf{T}\|_F$$

Now we calculate $\|\mathbf{A}\mathbf{A}^\mathsf{T}\|_F$

Start with the fact that

$$\|\mathbf{A}\mathbf{A}^\mathsf{T}\|_F^2 = \mathrm{tr}\left\{\mathbf{A}\mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{A}^\mathsf{T}\right\}$$

Small Coherence

How can we make the coherence small?

We know that $\|\mathbf{I}_M\|_F^2 = M$*; thus,*

$$\mathrm{tr}\left\{\mathbf{A}\mathbf{A}^\mathsf{T}\right\} \leq \sqrt{M}\|\mathbf{A}\mathbf{A}^\mathsf{T}\|_F$$

Now we calculate $\|\mathbf{A}\mathbf{A}^\mathsf{T}\|_F$

Now, we circularly rotate

$$\|\mathbf{A}\mathbf{A}^\mathsf{T}\|_F^2 = \mathrm{tr}\left\{\mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{A}^\mathsf{T}\mathbf{A}\right\}$$

Small Coherence

How can we make the coherence small?

> *We know that* $\|\mathbf{I}_M\|_F^2 = M$; *thus,*
> $$\mathrm{tr}\left\{\mathbf{A}\mathbf{A}^\mathsf{T}\right\} \leq \sqrt{M}\|\mathbf{A}\mathbf{A}^\mathsf{T}\|_F$$

Now we calculate $\|\mathbf{A}\mathbf{A}^\mathsf{T}\|_F$

> *Now, we we define* $\mathbf{G} = \mathbf{A}^\mathsf{T}\mathbf{A}$
> $$\|\mathbf{A}\mathbf{A}^\mathsf{T}\|_F^2 = \mathrm{tr}\left\{\mathbf{G}\mathbf{G}^\mathsf{T}\right\}$$

Small Coherence

How can we make the coherence small?

So, we could write

$$\operatorname{tr}\left\{\mathbf{A}\mathbf{A}^{\mathsf{T}}\right\} \leq \sqrt{M}\sqrt{\operatorname{tr}\left\{\mathbf{G}\mathbf{G}^{\mathsf{T}}\right\}}$$

Or equivalently

$$\operatorname{tr}\left\{\mathbf{A}\mathbf{A}^{\mathsf{T}}\right\}^2 \leq M\operatorname{tr}\left\{\mathbf{G}\mathbf{G}^{\mathsf{T}}\right\}$$

Let's now calculate every component

Small Coherence

How can we make the coherence small?

$$\mathrm{tr}\left\{\mathbf{A}\mathbf{A}^{\mathsf{T}}\right\}^2 \leq M \mathrm{tr}\left\{\mathbf{G}\mathbf{G}^{\mathsf{T}}\right\}$$

We calculate the components now

By circular shift, we have

$$\mathrm{tr}\left\{\mathbf{A}\mathbf{A}^{\mathsf{T}}\right\} = \mathrm{tr}\left\{\mathbf{A}^{\mathsf{T}}\mathbf{A}\right\}$$

Let's get back to representation $\mathbf{A} = [\boldsymbol{a}_1, \dots, \boldsymbol{a}_N]$

Small Coherence

How can we make the coherence small?

$$\mathrm{tr}\left\{\mathbf{A}\mathbf{A}^{\mathsf{T}}\right\}^2 \leq M\mathrm{tr}\left\{\mathbf{G}\mathbf{G}^{\mathsf{T}}\right\}$$

We calculate the components now

By the representation, we could say

$$\mathrm{tr}\left\{\mathbf{A}^{\mathsf{T}}\mathbf{A}\right\} = \sum_{n=1}^{N} \|\boldsymbol{a}_n\|^2 = N$$

But, we know that $\|\boldsymbol{a}_n\|^2 = 1$

Small Coherence

How can we make the coherence small?

$$N^2 \leq M \mathrm{tr} \left\{ \mathbf{G}\mathbf{G}^\mathsf{T} \right\}$$

We calculate the components now

Now, consider

$$\mathbf{G} = \mathbf{A}^\mathsf{T}\mathbf{A}$$

It is $N \times N$ matrix whose entries are $[\mathbf{G}]_{i,j} = \boldsymbol{a}_i^\mathsf{T}\boldsymbol{a}_j$

Small Coherence

How can we make the coherence small?

$$N^2 \leq M \operatorname{tr}\left\{\mathbf{G}\mathbf{G}^{\mathsf{T}}\right\}$$

We calculate the components now

So, we could say

$$\begin{aligned}\operatorname{tr}\left\{\mathbf{G}\mathbf{G}^{\mathsf{T}}\right\} &= \|\mathbf{G}\|_F^2 \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N}\left|\boldsymbol{a}_i^{\mathsf{T}}\boldsymbol{a}_j\right|^2\end{aligned}$$

Small Coherence

How can we make the coherence small?

$$N^2 \leq M \operatorname{tr}\left\{ \mathbf{G}\mathbf{G}^\mathsf{T} \right\}$$

We calculate the components now

Well, we now extend this term

$$\|\mathbf{G}\|_F^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} \left| \boldsymbol{a}_i^\mathsf{T} \boldsymbol{a}_j \right|^2 = \sum_{i=1}^{N} \left| \boldsymbol{a}_i^\mathsf{T} \boldsymbol{a}_i \right|^2 + \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \left| \boldsymbol{a}_i^\mathsf{T} \boldsymbol{a}_j \right|^2$$

$$= \sum_{i=1}^{N} \|\boldsymbol{a}_i\|^2 + \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \left| \boldsymbol{a}_i^\mathsf{T} \boldsymbol{a}_j \right|^2$$

Small Coherence

How can we make the coherence small?

$$N^2 \leq M \mathrm{tr}\left\{ \mathbf{G}\mathbf{G}^\mathsf{T} \right\}$$

We calculate the components now

Well, we now extend this term

$$\|\mathbf{G}\|_F^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} \left| \boldsymbol{a}_i^\mathsf{T} \boldsymbol{a}_j \right|^2 = \sum_{i=1}^{N} \left| \boldsymbol{a}_i^\mathsf{T} \boldsymbol{a}_i \right|^2 + \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \left| \boldsymbol{a}_i^\mathsf{T} \boldsymbol{a}_j \right|^2$$

$$= N + \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \left| \boldsymbol{a}_i^\mathsf{T} \boldsymbol{a}_j \right|^2$$

Small Coherence

How can we make the coherence small?

$$N^2 \leq M \mathrm{tr} \left\{ \mathbf{G}\mathbf{G}^\mathsf{T} \right\}$$

We calculate the components now

So, we have

$$\mathrm{tr} \left\{ \mathbf{G}\mathbf{G}^\mathsf{T} \right\} = \|\mathbf{G}\|_F^2 = N + \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \left| \boldsymbol{a}_i^\mathsf{T} \boldsymbol{a}_j \right|^2$$

By definition of coherence: $\left| \boldsymbol{a}_i^\mathsf{T} \boldsymbol{a}_j \right| \leq \mu\left(\mathbf{A}\right)$ *for any* $j \neq i$

Small Coherence

How can we make the coherence small?

$$N^2 \leq M \operatorname{tr}\left\{\mathbf{G}\mathbf{G}^{\mathsf{T}}\right\}$$

We calculate the components now

So, we could say

$$\begin{aligned}\operatorname{tr}\left\{\mathbf{G}\mathbf{G}^{\mathsf{T}}\right\} = \|\mathbf{G}\|_F^2 &= N + \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \left|\boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_j\right|^2 \\ &\leq N + N(N-1)\,\mu^2(\mathbf{A})\end{aligned}$$

Small Coherence

How can we make the coherence small?

So, we could finally say

$$N^2 \leq M \mathrm{tr}\left\{\mathbf{G}\mathbf{G}^{\mathsf{T}}\right\} \leq M\left(N + N\left(N-1\right)\mu^2\left(\mathbf{A}\right)\right)$$

We do we conclude then?

For any matrix, we have

$$\mu\left(\mathbf{A}\right) \geq \sqrt{\frac{N-M}{M\left(N-1\right)}}$$

Small Coherence

How can we make the coherence small?

Now, we can answer this question

$$\mu(\mathbf{A}) \geq \sqrt{\frac{N - M}{M(N-1)}}$$

To reduce $\mu(\mathbf{A})$, we should increase M

We need enough samples to make sure, we can recover

This is consistent with our initial intuition

Small Coherence

Attention! *The inequality*

$$\mu(\mathbf{A}) \geq \sqrt{\frac{N - M}{M(N-1)}}$$

becomes an identity, if

There exists a constant ρ, such that

$$\mathbf{A}\mathbf{A}^{\mathsf{T}} = \rho \mathbf{I}_M$$

It means that $\mathbf{A}$ *constructs a tight frame*

Small Coherence

Attention! *The inequality*

$$\mu(\mathbf{A}) \geq \sqrt{\frac{N-M}{M(N-1)}}$$

becomes an identity, if

There exists a constant α, such that for all $i \neq j$

$$\boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_j = \alpha$$

It means that $\mathbf{A}$ *constructs an equiangular frame*

Small Coherence

Attention! *The inequality*

$$\mu(\mathbf{A}) \geq \sqrt{\frac{N - M}{M(N-1)}}$$

becomes an identity, if

$\mathbf{A}$ *constructs an equiangular tight frame*

Small Coherence

This result extends to ℓ_1 coherence function

Lower Bound on ℓ_1 Coherence Function

For $\mathbf{A} \in \mathbb{R}^{M \times N}$, we have

$$\mu_1(s) \geq s\sqrt{\frac{N-M}{M(N-1)}}$$

if $s < \sqrt{N-1}$

The bound is tight if

A *constructs an equiangular tight frame*

Coherence of Matrices

Section 3: Recovery Guarantee for OMP

Back to Our Initial Example

We considered the OMP algorithm in noise-free case

- *Start with* $\mathbf{x}^{(0)} = \mathbf{0}$ *and* $\mathcal{S}^{(0)} = \emptyset$
- *Update the support for* $t \geq 1$ *as*

$$\mathcal{S}^{(t)} = \mathcal{S}^{(t-1)} \cup \operatorname*{argmax}_{n} \left| \left[\mathbf{A}^{\mathsf{T}} \left(\mathbf{y} - \mathbf{A}\mathbf{x}^{(t-1)} \right) \right]_n \right|$$

- *Set* $\mathbf{z} = \mathbf{A}^{\dagger}_{\mathcal{S}^{(t)}} \mathbf{y}$ *and update* $\mathbf{x}^{(t)}$ *as*

$$\mathbf{x}^{(t)}_{\mathcal{S}^{(t)}} = \mathbf{z}$$

- *Stop at iteration* T, *at which* $\mathbf{A}\mathbf{x}^{(T)} = \mathbf{y}$

Back to Our Initial Example

After some derivations, we concluded that ...

Recovery Guarantee for OMP

If for every $\mathcal{S} \subset \{1, \dots, N\}$ *with* $|\mathcal{S}| = s$*, we have*

1. $\mathbf{A}_{\mathcal{S}}$ *is injective*
2. *For all* $\boldsymbol{r} \in \mathcal{U} = \{\boldsymbol{r} = \mathbf{A}\boldsymbol{z} : \mathrm{Supp}(\boldsymbol{z}) \subseteq \mathcal{S}\}$, $\mathbf{A}$ *reads*

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right| > \max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right|$$

Then, OMP recovers any s-sparse signal sampled by $\mathbf{A}$

Back to Our Initial Example

What was the downside of this result?

The constraint is tough to be checked!

Could we do something about this now?

Well, as we learned coherence, we try to process it, such that

We conclude a constraint in terms of coherence

Recovery Guarantee for OMP

Let's start with the first constraint

1. $\mathbf{A}_{\mathcal{S}}$ *is injective*

We learned in the previous section that

A Result of Bounding Spectrum of Matrices

For any $\mathcal{S} \subset \{1, \dots, M\}$ *with* $|\mathcal{S}| = s$, $\mathbf{A}_{\mathcal{S}}$ *is injective if*

$$1 - \mu_1 (s - 1) > 0$$

Recovery Guarantee for OMP

Let's start with the first constraint

$\boxed{1}$ $\mathbf{A}_{\mathcal{S}}$ *is injective*

Since $\mathbf{A}$ *is a fat matrix, we know that*

$$\mu_1(s) \neq 0 \rightsquigarrow \mu_1(s) > 0$$

So, we could say if

$$1 - \mu_1(s-1) > \mu_1(s) \rightsquigarrow 1 - \mu_1(s-1) > 0$$

Recovery Guarantee for OMP

Let's start with the first constraint

1 $\mathbf{A}_{\mathcal{S}}$ *is injective*

So, we could conclude this simple sufficient constraint

$\mathbf{A}_{\mathcal{S}}$ *is injective, if*

$$\mu_1(s) + \mu_1(s-1) < 1$$

Keep this in mind!

Recovery Guarantee for OMP

Now, let's play a bit with the second constraint

2. *For all* $\boldsymbol{r} \in \mathcal{U} = \{\boldsymbol{r} = \mathbf{A}\boldsymbol{z} : \operatorname{Supp}(\boldsymbol{z}) \subseteq \mathcal{S}\}$, $\mathbf{A}$ *reads*

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right| > \max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right|$$

Assume we focus on a $\boldsymbol{r}$

There is a $\boldsymbol{z}$ *with* $\operatorname{Supp}(\boldsymbol{z}) = \mathcal{S}$, *such that*

$$\boldsymbol{r} = \mathbf{A}\boldsymbol{z} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N] \begin{bmatrix} z_1 \\ \vdots \\ z_N \end{bmatrix} = \sum_{i \in \mathcal{S}} z_i \boldsymbol{a}_i$$

Recovery Guarantee for OMP

Now, let's play a bit with the second constraint

2 *For all* $\mathbf{r} \in \mathcal{U} = \{\mathbf{r} = \mathbf{A}\mathbf{z} : \mathrm{Supp}(\mathbf{z}) \subseteq \mathcal{S}\}$, $\mathbf{A}$ *reads*

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \mathbf{r} \right]_n \right| > \max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \mathbf{r} \right]_n \right|$$

Assume we focus on an $\mathbf{r}$

We define $i^\star$ *to be*

$$i^\star = \max_{i \in \mathcal{S}} |z_i| = \max_{1 \leq i \leq N} |z_i|$$

Recovery Guarantee for OMP

Now, let's play a bit with the second constraint

2 *For all* $\boldsymbol{r} \in \mathcal{U} = \{\boldsymbol{r} = \mathbf{A}\boldsymbol{z} : \operatorname{Supp}(\boldsymbol{z}) \subseteq \mathcal{S}\}$, $\mathbf{A}$ *reads*

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right| > \max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right|$$

Although this constraint is tough to check, we try to find

Bounds B_1 *and* B_2, *such that*

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right| \geq B_1 > B_2 \geq \max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right|$$

Recovery Guarantee for OMP

Bounds B_1 and B_2, such that

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right| \geq B_1 > B_2 \geq \max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right|$$

Consider an index $n \notin \mathcal{S}$, we can write

$$\left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n = \left[\begin{bmatrix} \boldsymbol{a}_1^\mathsf{T} \\ \vdots \\ \boldsymbol{a}_N^\mathsf{T} \end{bmatrix} \boldsymbol{r} \right]_n = \boldsymbol{a}_n^\mathsf{T} \boldsymbol{r} = \boldsymbol{a}_n^\mathsf{T} \sum_{i \in \mathcal{S}} z_i \boldsymbol{a}_i = \sum_{i \in \mathcal{S}} z_i \boldsymbol{a}_n^\mathsf{T} \boldsymbol{a}_i$$

Recovery Guarantee for OMP

Bounds B_1 and B_2, such that

$$\max_{n\in\mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right| \geq B_1 > B_2 \geq \max_{n\notin\mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right|$$

So, we could say for $n \notin \mathcal{S}$

$$\left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right| = \left| \sum_{i\in\mathcal{S}} z_i \boldsymbol{a}_n^\mathsf{T} \boldsymbol{a}_i \right| \leq \sum_{i\in\mathcal{S}} \left| z_i \boldsymbol{a}_n^\mathsf{T} \boldsymbol{a}_i \right|$$
$$= \sum_{i\in\mathcal{S}} |z_i| \left| \boldsymbol{a}_n^\mathsf{T} \boldsymbol{a}_i \right|$$

Recovery Guarantee for OMP

Bounds B_1 and B_2, such that

$$\max_{n\in\mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right| \geq B_1 > B_2 \geq \max_{n\notin\mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right|$$

Since we have $|z_{i^\star}| \geq |z_i|$ for any i, we can further write

$$\begin{aligned} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right| &\leq \sum_{i\in\mathcal{S}} |z_i| \left| \boldsymbol{a}_n^{\mathsf{T}} \boldsymbol{a}_i \right| \leq \sum_{i\in\mathcal{S}} |z_{i^\star}| \left| \boldsymbol{a}_n^{\mathsf{T}} \boldsymbol{a}_i \right| \\ &= |z_{i^\star}| \sum_{i\in\mathcal{S}} \left| \boldsymbol{a}_n^{\mathsf{T}} \boldsymbol{a}_i \right| \end{aligned}$$

Recovery Guarantee for OMP

Bounds B_1 and B_2, such that

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right| \geq B_1 > B_2 \geq \max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right|$$

Remember the definition of sum-correlation

$$\begin{aligned} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right| &\leq |z_{i^\star}| \sum_{i \in \mathcal{S}} \left| \boldsymbol{a}_n^{\mathsf{T}} \boldsymbol{a}_i \right| \\ &= |z_{i^\star}| \, \sigma_n (\mathbf{A})_{\mathcal{S}} \leq |z_{i^\star}| \, \mu_1 (s) \end{aligned}$$

Remember that $\sigma_n (\mathbf{A})_{\mathcal{S}} \leq \mu_1 (s)$ for any $|\mathcal{S}| = s$ and any $n \notin \mathcal{S}$

Recovery Guarantee for OMP

Bounds B_1 and B_2, such that

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right| \geq B_1 > B_2 \geq \max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right|$$

Since it holds for any $n \notin \mathcal{S}$, we could conclude that

$$\max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right| \leq |z_{i^\star}| \, \mu_1 (s)$$

So, a possible choice for B_2 is

$$B_2 = |z_{i^\star}| \, \mu_1 (s)$$

Recovery Guarantee for OMP

Bounds B_1 and B_2, such that

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right| \geq B_1 > B_2 \geq \max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right|$$

Now, we focus on $\boldsymbol{a}_{i^\star}$: We know that

$$\left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_{i^\star} \right| = \left| \boldsymbol{a}_{i^\star}^\mathsf{T} \boldsymbol{r} \right| = \left| \sum_{i \in \mathcal{S}} z_i \boldsymbol{a}_{i^\star}^\mathsf{T} \boldsymbol{a}_i \right|$$

$$= \left| z_{i^\star} \boldsymbol{a}_{i^\star}^\mathsf{T} \boldsymbol{a}_{i^\star} + \sum_{i^\star \neq i \in \mathcal{S}} z_i \boldsymbol{a}_{i^\star}^\mathsf{T} \boldsymbol{a}_i \right|$$

Recovery Guarantee for OMP

Bounds B_1 and B_2, such that

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right| \geq B_1 > B_2 \geq \max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right|$$

Using triangle inequality, we have

$$\left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_{i^\star} \right| = \left| z_{i^\star} \underbrace{\| \boldsymbol{a}_{i^\star} \|^2}_{1} + \sum_{i^\star \neq i \in \mathcal{S}} z_i \boldsymbol{a}_{i^\star}^{\mathsf{T}} \boldsymbol{a}_i \right|$$

$$\geq |z_{i^\star}| - \left| \sum_{i^\star \neq i \in \mathcal{S}} z_i \boldsymbol{a}_{i^\star}^{\mathsf{T}} \boldsymbol{a}_i \right|$$

Recovery Guarantee for OMP

Bounds B_1 and B_2, such that

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right| \geq B_1 > B_2 \geq \max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right|$$

Again by triangle inequality, we have

$$
\begin{aligned}
-\left| \sum_{i^\star \neq i \in \mathcal{S}} z_i \boldsymbol{a}_{i^\star}^\mathsf{T} \boldsymbol{a}_i \right| &\geq - \sum_{i^\star \neq i \in \mathcal{S}} \left| z_i \boldsymbol{a}_{i^\star}^\mathsf{T} \boldsymbol{a}_i \right| \\
&\geq - \left| z_{i^\star} \right| \sum_{i^\star \neq i \in \mathcal{S}} \left| \boldsymbol{a}_{i^\star}^\mathsf{T} \boldsymbol{a}_i \right| = - \left| z_{i^\star} \right| \sigma_{i^\star} \left(\mathbf{A} \right)_{\mathcal{S} - \{ i^\star \}}
\end{aligned}
$$

Recovery Guarantee for OMP

Bounds B_1 and B_2, such that

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right| \geq B_1 > B_2 \geq \max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right|$$

By definition, we could further say that

$$-\left| \sum_{i^\star \neq i \in \mathcal{S}} z_i \boldsymbol{a}_{i^\star}^\mathsf{T} \boldsymbol{a}_i \right| \geq -\left| z_{i^\star} \right| \sigma_{i^\star} \left(\mathbf{A} \right)_{\mathcal{S} - \{i^\star\}} \geq -\left| z_{i^\star} \right| \mu_1 \left(s - 1 \right)$$

Note that $\left| \mathcal{S} - \{ i^\star \} \right| = s - 1$ *and* $i^\star \notin \mathcal{S} - \{ i^\star \}$

Recovery Guarantee for OMP

Bounds B_1 and B_2, such that

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right| \geq B_1 > B_2 \geq \max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right|$$

Back to the lower bound, we have

$$\begin{aligned} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_{i^\star} \right| &\geq \left| z_{i^\star} \right| - \left| \sum_{i^\star \neq i \in \mathcal{S}} z_i \boldsymbol{a}_{i^\star}^\mathsf{T} \boldsymbol{a}_i \right| \\ &\geq \left| z_{i^\star} \right| - \left| z_{i^\star} \right| \mu_1 \left(s - 1 \right) \end{aligned}$$

Recovery Guarantee for OMP

Bounds B_1 and B_2, such that

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right| \geq B_1 > B_2 \geq \max_{n \notin \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right|$$

Since $i^\star \in \mathcal{S}$, we could say

$$\max_{n \in \mathcal{S}} \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_n \right| \geq \left| \left[\mathbf{A}^\mathsf{T} \boldsymbol{r} \right]_{i^\star} \right| \geq \left| z_{i^\star} \right| - \left| z_{i^\star} \right| \mu_1 \left(s - 1 \right)$$

So, a possible choice for B_1 is

$$B_1 = \left| z_{i^\star} \right| - \left| z_{i^\star} \right| \mu_1 \left(s - 1 \right)$$

Recovery Guarantee for OMP

Bounds B_1 and B_2, such that

$$\max_{n\in\mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right| \geq B_1 > B_2 \geq \max_{n\notin\mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right|$$

For the given choices of B_1 and B_2, the sufficient condition is

$$|z_{i^\star}| - |z_{i^\star}| \, \mu_1 (s-1) > |z_{i^\star}| \, \mu_1 (s)$$

Or equivalently

$$1 - \mu_1(s-1) > \mu_1(s) \rightsquigarrow \mu_1(s) + \mu_1(s-1) < 1$$

Recovery Guarantee for OMP

What do we conclude?

If we have

$$\mu_1(s) + \mu_1(s-1) < 1$$

Then, we are sure that

1. $\mathbf{A}_{\mathcal{S}}$ *is injective*
2. *For all* $\boldsymbol{r} \in \mathcal{U} = \{\boldsymbol{r} = \mathbf{A}\boldsymbol{z} : \mathrm{Supp}(\boldsymbol{z}) \subseteq \mathcal{S}\}$, $\mathbf{A}$ *reads*

$$\max_{n\in\mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right| > \max_{n\notin\mathcal{S}} \left| \left[\mathbf{A}^{\mathsf{T}} \boldsymbol{r} \right]_n \right|$$

Recovery Guarantee for OMP

Recovery Guarantee for OMP

If the sampling matrix $\mathbf{A}$ *satisfies*

$$\mu_1(s) + \mu_1(s-1) < 1$$

Then, the OMP recovers any s-sparse signal from the noise-free samples collected by $\mathbf{A}$ *after s iterations*

Is this constraint computationally simple to check?

No! But we could use the upper bound

Recovery Guarantee for OMP

Recovery Guarantee for OMP

If the sampling matrix $\mathbf{A}$ *satisfies*

$$\mu_1(s) + \mu_1(s-1) < 1$$

Then, the OMP recovers any s-sparse signal from the noise-free samples collected by $\mathbf{A}$ *after s iterations*

Using the upper bound $\mu_1(s) \leq s\mu(\mathbf{A})$, *we have*

$$\mu_1(s) + \mu_1(s-1) < (2s-1)\,\mu(\mathbf{A})$$

Recovery Guarantee for OMP

Alternative Recovery Guarantee for OMP

If the sampling matrix $\mathbf{A}$ *satisfies*

$$\mu(\mathbf{A}) < \frac{1}{2s-1}$$

Then, the OMP recovers any s-sparse signal from the noise-free samples collected by $\mathbf{A}$ *after s iterations*

We could achieve this only if

$$M > \frac{N(2s-1)^2}{(2s-1)^2 + N - 1}$$

Recovery Guarantee for OMP

Example: *For an $s = 0.05N$-sparse signal, optimal recovery needs*

$$M \approx 2s = 0.1N$$

samples, but the OMP needs

$$M \approx \frac{N\left(2s-1\right)^2}{\left(2s-1\right)^2 + N - 1}$$

This means that when $N = 100$,

- *OMP needs $M \approx 45$ and optimal recovery needs $M \approx 10$*

Poor performance at the expense of low complexity?!

Coherence of Matrices

Section 4: Recovery Guarantee for Basis Pursuit

Our Initial Results on Basis Pursuit

For the noise-free case, we showed that

Recovery Guarantee by Null Space Property

Every s-sparse signal $\boldsymbol{x}$ is uniquely recovered via basis pursuit from its samples collected by sampling matrix $\mathbf{A}$, *if and only if* $\mathbf{A}$ *satisfies ℓ_1 null space property of order s*

This means that basis pursuit uniquely recovers $\boldsymbol{x}$ from $\boldsymbol{y} = \mathbf{A}\boldsymbol{x}$, if

$\mathbf{A}$ *satisfies the null space property!*

Our Initial Results on Basis Pursuit

$\boxed{\mathbf{A} \textit{ satisfies the null space property!}}$

This means that for every subset $\mathcal{S}$ *with* $|\mathcal{S}| = s$

$$\|\boldsymbol{v}_{\mathcal{S}}\|_1 < \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1$$

for every $\boldsymbol{v} \in \ker \mathbf{A}$

We now check, if we could find a constraint on $\mathbf{A}$*, such that*

The fulfillment of null space property is guaranteed

Recovery Guarantee for Basis Pursuit

Let's focus on a support $\mathcal{S}$ with $|\mathcal{S}| = s$

For every $\boldsymbol{v} \in \ker \mathbf{A}$, we have

$$\mathbf{0} = \mathbf{A}\boldsymbol{v} = [\boldsymbol{a}_1, \dots, \boldsymbol{a}_N] \begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix}$$

$$= \sum_{n=1}^{N} v_n \boldsymbol{a}_n$$

Recovery Guarantee for Basis Pursuit

Let's now choose an index $i \in \mathcal{S}$

For every $\boldsymbol{v} \in \ker \mathbf{A}$, we have

$$0 = \boldsymbol{a}_i^{\mathsf{T}} \mathbf{0} = \boldsymbol{a}_i^{\mathsf{T}} \sum_{n=1}^{N} v_n \boldsymbol{a}_n$$

$$= \sum_{n=1}^{N} v_n \boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_n$$

$$= v_i \underbrace{\boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_i}_{1} + \sum_{n \in \mathcal{S}, n \neq i} v_n \boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_n + \sum_{n \notin \mathcal{S}} v_n \boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_n$$

Recovery Guarantee for Basis Pursuit

Let's now choose an index $i \in \mathcal{S}$

For every $\boldsymbol{v} \in \ker \mathbf{A}$, we have

$$v_i = - \sum_{n \in \mathcal{S}, n \neq i} v_n \boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_n - \sum_{n \notin \mathcal{S}} v_n \boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_n$$

So, we could say that for any $i \in \mathcal{S}$

$$\begin{aligned} |v_i| &= \left| \sum_{n \in \mathcal{S}, n \neq i} v_n \boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_n + \sum_{n \notin \mathcal{S}} v_n \boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_n \right| \\ &\leq \sum_{n \in \mathcal{S}, n \neq i} |v_n| \left| \boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_n \right| + \sum_{n \notin \mathcal{S}} |v_n| \left| \boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_n \right| \end{aligned}$$

Recovery Guarantee for Basis Pursuit

$$\boxed{\left|v_i\right| \leq \sum_{n \in \mathcal{S}, n \neq i} \left|v_n\right| \left|\boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_n\right| + \sum_{n \notin \mathcal{S}} \left|v_n\right| \left|\boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_n\right|}$$

Let's now sum it over $i \in \mathcal{S}$

$$\begin{aligned}\sum_{i \in \mathcal{S}} \left|v_i\right| &\leq \sum_{n \in \mathcal{S}} \left|v_n\right| \sum_{i \in \mathcal{S}, i \neq n} \left|\boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_n\right| + \sum_{n \notin \mathcal{S}} \left|v_n\right| \sum_{i \in \mathcal{S}} \left|\boldsymbol{a}_i^{\mathsf{T}} \boldsymbol{a}_n\right| \\ &= \sum_{n \in \mathcal{S}} \left|v_n\right| \sigma_n \left(\mathbf{A}\right)_{\mathcal{S}-\{n\}} + \sum_{n \notin \mathcal{S}} \left|v_n\right| \sigma_n \left(\mathbf{A}\right)_{\mathcal{S}}\end{aligned}$$

Recovery Guarantee for Basis Pursuit

We now use the fact that

$$\sigma_n(\mathbf{A})_{\mathcal{S}-\{n\}} \leq \mu_1(s-1) \qquad \sigma_n(\mathbf{A})_{\mathcal{S}} \leq \mu_1(s)$$

This means

$$\begin{aligned}\sum_{i\in\mathcal{S}} |v_i| &\leq \sum_{n\in\mathcal{S}} |v_n| \sigma_n(\mathbf{A})_{\mathcal{S}-\{n\}} + \sum_{n\notin\mathcal{S}} |v_n| \sigma_n(\mathbf{A})_{\mathcal{S}} \\ &\leq \sum_{n\in\mathcal{S}} |v_n| \mu_1(s-1) + \sum_{n\notin\mathcal{S}} |v_n| \mu_1(s)\end{aligned}$$

Recovery Guarantee for Basis Pursuit

We now note that

$$\sum_{i \in \mathcal{S}} |v_i| = \sum_{n \in \mathcal{S}} |v_n| = \|\boldsymbol{v}_{\mathcal{S}}\|_1$$

$$\sum_{n \notin \mathcal{S}} |v_n| = \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1$$

Thus, we could conclude that

$$\|\boldsymbol{v}_{\mathcal{S}}\|_1 \leq \mu_1 (s-1) \|\boldsymbol{v}_{\mathcal{S}}\|_1 + \mu_1 (s) \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1$$

Recovery Guarantee for Basis Pursuit

We now note that

$$\sum_{i \in \mathcal{S}} |v_i| = \sum_{n \in \mathcal{S}} |v_n| = \|\boldsymbol{v}_{\mathcal{S}}\|_1$$

$$\sum_{n \notin \mathcal{S}} |v_n| = \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1$$

Or equivalently

$$(1 - \mu_1 (s - 1)) \|\boldsymbol{v}_{\mathcal{S}}\|_1 \leq \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1 \mu_1 (s)$$

Recovery Guarantee for Basis Pursuit

What have we concluded up to now?

> *For any support* $\mathcal{S}$ *with* $|\mathcal{S}| = s$, *any* $\boldsymbol{v} \in \ker \mathbf{A}$ *satisfy*
>
> $$(1 - \mu_1(s-1)) \|\boldsymbol{v}_{\mathcal{S}}\|_1 \leq \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1 \mu_1(s)$$

If we know that

$$1 - \mu_1(s-1) > \mu_1(s)$$

Then, we could guarantee for sure that

$$\|\boldsymbol{v}_{\mathcal{S}}\|_1 < \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1$$

Recovery Guarantee for Basis Pursuit

Sufficient Condition for Null Space Property

If matrix $\mathbf{A}$ *satisfies*

$$\mu_1(s-1) + \mu_1(s) < 1$$

Then, it satisfies the null space property of order s

So we had

$$\mu_1(s-1) + \mu_1(s) < 1 \rightsquigarrow \textit{Null Space Property}$$

and now, we have

$$\textit{Null Space Property} \rightsquigarrow \textit{Recovery Guarantee}$$

Recovery Guarantee for Basis Pursuit

Recovery Guarantee for Basis Pursuit

Every s-sparse signal $\boldsymbol{x}$ is uniquely recovered via basis pursuit from its samples collected by sampling matrix $\mathbf{A}$, if

$$\mu_1(s-1) + \mu_1(s) < 1$$

Coherence of Matrices

Section 5: Recovery Guarantee for Thresholding Algorithms

Iterative Hard Thresholding

Remember iterative hard thresholding in noise-free scenario

- *It starts with* $\boldsymbol{x}^{(0)} = \mathbf{0}$ *and* $\mathcal{S}^{(0)} = \emptyset$
- *It updates the approximation of the signal as*

$$\boldsymbol{x}^{(t)} = T_s^{\mathrm{H}} \left(\boldsymbol{x}^{(t-1)} + \mathbf{A}^{\mathsf{T}} \left(\boldsymbol{y} - \mathbf{A}\boldsymbol{x}^{(t-1)} \right) \right)$$

- *It stops at iteration* T, *at which* $\mathbf{A}\boldsymbol{x}^{(T)} = \boldsymbol{y}$

Remember that $T_s^{\mathrm{H}}(\boldsymbol{x})$ *keep only* s *largest entries of* $\boldsymbol{x}$

Iterative Hard Thresholding

Using the similar approach, we can show that

If matrix $\mathbf{A}$ *satisfies*

$$\mu_1 (s - 1) + 2\mu_1 (s) < 1$$

Then, the t largest value of

$$\boldsymbol{u}^{(t)} = \boldsymbol{x}^{(t-1)} + \mathbf{A}^{\mathsf{T}} \left(\boldsymbol{y} - \mathbf{A}\boldsymbol{x}^{(t-1)} \right)$$

contains the t largest value of $\boldsymbol{x}$

This is shown simply by induction

Iterative Hard Thresholding

What does that mean?

If matrix $\mathbf{A}$ *satisfies* $\mu_1(s-1) + 2\mu_1(s) < 1$,

$$\boldsymbol{x}^{(1)} = T_s^{\mathrm{H}}\left(\boldsymbol{x}^{(0)} + \mathbf{A}^{\mathsf{T}}\left(\boldsymbol{y} - \mathbf{A}\boldsymbol{x}^{(0)}\right)\right)$$

contains

- *the largest entry of* $\boldsymbol{x}$
- $s-1$ *other entries*
- $N-s$ *zeros*

Iterative Hard Thresholding

What does that mean?

If matrix $\mathbf{A}$ *satisfies* $\mu_1(s-1) + 2\mu_1(s) < 1$,

$$\boldsymbol{x}^{(2)} = T_s^{\mathrm{H}}\left(\boldsymbol{x}^{(1)} + \mathbf{A}^{\mathsf{T}}\left(\boldsymbol{y} - \mathbf{A}\boldsymbol{x}^{(1)}\right)\right)$$

contains

- *the two largest entry of* $\boldsymbol{x}$
- $s-2$ *other entries*
- $N - s$ *zeros*

Iterative Hard Thresholding

What does that mean?

If matrix $\mathbf{A}$ *satisfies* $\mu_1(s-1) + 2\mu_1(s) < 1$,

$$\boldsymbol{x}^{(s)} = T_s^{\mathrm{H}}\left(\boldsymbol{x}^{(s-1)} + \mathbf{A}^{\mathsf{T}}\left(\boldsymbol{y} - \mathbf{A}\boldsymbol{x}^{(s-1)}\right)\right)$$

contains

- *the* s *largest entry of* $\boldsymbol{x}$
- $N - s$ *zeros*

The algorithm recovers $\boldsymbol{x}$*!*

Recovery Guarantee for Hard Thresholding

Recovery Guarantee for Hard Thresholding

Every s-sparse signal $\mathbf{x}$ is recovered via hard thresholding algorithm after at most s iterations from its samples collected by sampling matrix $\mathbf{A}$, if

$$\mu_1 (s-1) + 2\mu_1 (s) < 1$$

Coherence of Matrices

Section 6: Sufficiency Order of Coherence

Number of Sufficient Samples

Remember the lower bound on ℓ_1 coherence function

Lower Bound on ℓ_1 Coherence Function

For $\mathbf{A} \in \mathbb{R}^{M \times N}$, we have

$$\mu_1(s) \geq s \sqrt{\frac{N - M}{M(N-1)}}$$

if $s < \sqrt{N-1}$

The bound is tight if

A *constructs an equiangular tight frame*

Number of Sufficient Samples

Let's for the moment assume we can make an **A**, *such that*

$$\boxed{\textbf{A} \text{ constructs an equiangular tight frame}}$$

Then, to make sure that

either basis pursuit or OMP recovers any s-sparse signal

we need to have

$$\mu_1 (s-1) + \mu_1 (s) < 1$$

Number of Sufficient Samples

We now replace $\mu_1(\cdot)$ with its bound

$$(s-1)\sqrt{\frac{N-M}{M(N-1)}}+s\sqrt{\frac{N-M}{M(N-1)}}<1$$

Or equivalently, we need to have

$$(2s-1)\sqrt{\frac{N-M}{M(N-1)}}<1$$

Number of Sufficient Samples

Now, let us assume that $N \gg M$ and s, so we could write

$$\sqrt{\frac{N-M}{M\left(N-1\right)}} \approx \lim_{N\to\infty} \sqrt{\frac{N-M}{M\left(N-1\right)}}$$

while M is treated as a constant. So, we have

$$\sqrt{\frac{N-M}{M\left(N-1\right)}} \approx \frac{1}{\sqrt{M}}$$

Number of Sufficient Samples

We could more precisely write

$$\sqrt{\frac{N-M}{M(N-1)}} = \frac{C_{M,N}}{\sqrt{M}}$$

where $C_{M,N}$ is a bounded constant, i.e., it never gets very large

With this approximation, the number of sufficient samples reads

$$\frac{(2s-1)\, C_{M,N}}{\sqrt{M}} < 1$$

Number of Sufficient Samples

Note that $(2s - 1) < 2s$, *so we could say*

Coherence indicates that we need to have

$$\frac{(2C_{M,N}s)}{\sqrt{M}} \leq 1$$

Sufficiency Order of Coherence

Recovery guarantees based on the coherence of a matrix determine the sufficient number of samples for basis pursuit recovery as

$$M \geq Cs^2$$

for some constant C

Number of Sufficient Samples

Coherence calculates a quadratic growth in number of samples

But remember that our analysis of optimal approach said

In optimal case, we need

$$M \geq Cs$$

for some constant $1 < C \leq 2$

This is in fact linear!

Number of Sufficient Samples

Is it simply because basis pursuit is not optimal?

> *Well! Sub-optimally degrades the performance, but*
>
> *The predicted degradation in this case is huge!*

Then, what is the main reason for this huge degradation?

> *The prediction given by coherence is not tight!*

Number of Sufficient Samples

We hence go for the restricted isometry property

Restricted isometry property tells us that collecting

$$M \geq 2s \ln \left(\frac{N}{s} \right)$$

samples is sufficient to

recover an s-sparse signal with basis pursuit

Summary

In a nutshell, we have learned up to this point that

- *Null space property*
  - *Give necessary and sufficient conditions for recovery*
  - *It is computationally hard to be checked*
- *Alternatively, we can find guarantees in terms on coherence*
  - *We get rid of high complexity issue*
  - *Given results are only on sufficiency*

They are relatively loose bounds

We now learn the restricted isometry property

It gives tighter sufficient conditions

Restricted Isometry Property

Basic Definitions

We start with defining restricted isometry constant

Restricted Isometry Constant

The s-th restricted isometry constant of matrix $\mathbf{A}$ *is the minimum* $\delta \geq 0$ *such that for any s-sparse vector, we have*

$$(1-\delta)\left\|\boldsymbol{x}\right\|_2^2 \leq \left\|\mathbf{A}\boldsymbol{x}\right\|_2^2 \leq (1+\delta)\left\|\boldsymbol{x}\right\|_2^2$$

We show this constant with $\delta_s(\mathbf{A})$

What is the intuition behind this definition?

Let's play a bit with the definition

Basic Definitions

What is the intuition behind this definition?

> $\delta_s(\mathbf{A})$ *is the minimum value of* δ *for which we have*
>
> $$\left| \|\mathbf{A}\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2 \right| \leq \delta \|\boldsymbol{x}\|_2^2$$
>
> *when* $\boldsymbol{x}$ *is an* s*-sparse signal*

What happens when $\mathbf{A}$ *is orthogonal for any support* $|\mathcal{S}| = s$*?*

> *For any* s*-sparse signal, we have* $\mathbf{A}\boldsymbol{x} = \boldsymbol{x}$*. This mans that*
>
> $$\left| \|\mathbf{A}\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2 \right| = 0 \rightsquigarrow \delta_s(\mathbf{A}) = 0$$

Basic Definitions

What happens when $\mathbf{A}$ *is orthogonal for any support* $|\mathcal{S}| = s$*?*

$\boxed{\textit{This is in fact the ideal scenario!}}$

So, one could say

$\boxed{\delta_s(\mathbf{A}) \text{ calculates the deviation from this optimal case}}$

This means

$\boxed{\text{The smaller } \delta_s(\mathbf{A}) \text{ gets, the better } \mathbf{A} \text{ samples}}$

Basic Definitions

How can we calculate the restricted isometry constant?

$$\left| \|\mathbf{A}\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2 \right| \leq \delta_s\left(\mathbf{A}\right) \|\boldsymbol{x}\|_2^2$$

$\boldsymbol{x}$ *is an* s*-sparse signal; thus, for* $\mathcal{S} = \operatorname{Supp}\left(\boldsymbol{x}\right)$, *we get*

$$\begin{aligned} \left| \|\mathbf{A}_{\mathcal{S}}\boldsymbol{x}_{\mathcal{S}}\|_2^2 - \|\boldsymbol{x}_{\mathcal{S}}\|_2^2 \right| = \left| \|\mathbf{A}\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2 \right| &\leq \delta_s\left(\mathbf{A}\right) \|\boldsymbol{x}\|_2^2 \\ &= \delta_s\left(\mathbf{A}\right) \|\boldsymbol{x}_{\mathcal{S}}\|_2^2 \end{aligned}$$

So, this means for any $\mathcal{S}$

$$\left| \|\mathbf{A}_{\mathcal{S}}\boldsymbol{x}_{\mathcal{S}}\|_2^2 - \|\boldsymbol{x}_{\mathcal{S}}\|_2^2 \right| \leq \delta_s\left(\mathbf{A}\right) \|\boldsymbol{x}_{\mathcal{S}}\|_2^2$$

Basic Definitions

How can we calculate the restricted isometry constant?

So, we could say that

$$\frac{\left| \| \mathbf{A}_{\mathcal{S}} \boldsymbol{x}_{\mathcal{S}} \|_2^2 - \| \boldsymbol{x}_{\mathcal{S}} \|_2^2 \right|}{\| \boldsymbol{x}_{\mathcal{S}} \|_2^2} \leq \delta_s (\mathbf{A})$$

Or equivalently

$$\delta_s (\mathbf{A}) = \max_{\boldsymbol{x}_{\mathcal{S}} \neq \mathbf{0}, |\mathcal{S}| = s} \frac{\left| \| \mathbf{A}_{\mathcal{S}} \boldsymbol{x}_{\mathcal{S}} \|_2^2 - \| \boldsymbol{x}_{\mathcal{S}} \|_2^2 \right|}{\| \boldsymbol{x}_{\mathcal{S}} \|_2^2}$$

Basic Definitions

How can we calculate the restricted isometry constant?

Now, note that

$$\begin{aligned}\|\mathbf{A}_{\mathcal{S}}\boldsymbol{x}_{\mathcal{S}}\|_2^2 - \|\boldsymbol{x}_{\mathcal{S}}\|_2^2 &= \left(\mathbf{A}_{\mathcal{S}}\boldsymbol{x}_{\mathcal{S}}\right)^{\mathsf{T}}\mathbf{A}_{\mathcal{S}}\boldsymbol{x}_{\mathcal{S}} - \boldsymbol{x}_{\mathcal{S}}^{\mathsf{T}}\boldsymbol{x}_{\mathcal{S}} \\ &= \boldsymbol{x}_{\mathcal{S}}^{\mathsf{T}}\mathbf{A}_{\mathcal{S}}^{\mathsf{T}}\mathbf{A}_{\mathcal{S}}\boldsymbol{x}_{\mathcal{S}} - \boldsymbol{x}_{\mathcal{S}}^{\mathsf{T}}\mathbf{I}_s\boldsymbol{x}_{\mathcal{S}} \\ &= \boldsymbol{x}_{\mathcal{S}}^{\mathsf{T}}\left(\mathbf{A}_{\mathcal{S}}^{\mathsf{T}}\mathbf{A}_{\mathcal{S}} - \mathbf{I}_s\right)\boldsymbol{x}_{\mathcal{S}}\end{aligned}$$

Let could hence say that

$$\left|\|\mathbf{A}_{\mathcal{S}}\boldsymbol{x}_{\mathcal{S}}\|_2^2 - \|\boldsymbol{x}_{\mathcal{S}}\|_2^2\right| = \left|\boldsymbol{x}_{\mathcal{S}}^{\mathsf{T}}\left(\mathbf{A}_{\mathcal{S}}^{\mathsf{T}}\mathbf{A}_{\mathcal{S}} - \mathbf{I}_s\right)\boldsymbol{x}_{\mathcal{S}}\right|$$

Basic Definitions

Maximum $\mathbf{Q}$-Norm

For matrix $\mathbf{Q}$, *we have*

$$\left| \boldsymbol{x}^{\mathsf{T}} \mathbf{Q} \boldsymbol{x} \right| \leq \sigma_{\max} \left(\mathbf{Q} \right) \| \boldsymbol{x} \|^2$$

where $\sigma_{\max} \left(\mathbf{Q} \right)$ *is the maximum singular value*

$$\sigma_{\max} \left(\mathbf{Q} \right) = \sqrt{\lambda_{\max} \left(\mathbf{Q} \mathbf{Q}^{\mathsf{T}} \right)}$$

This bound is achievable, i.e. there exists $\boldsymbol{x}_{Opt}$ *such that*

$$\left| \boldsymbol{x}_{Opt}^{\mathsf{T}} \mathbf{Q} \boldsymbol{x}_{Opt} \right| = \sigma_{\max} \left(\mathbf{Q} \right) \| \boldsymbol{x}_{Opt} \|^2$$

Basic Definitions

How can we calculate the restricted isometry constant?

So for every $|\mathcal{S}| = s$, *we have*

$$\frac{\boldsymbol{x}_{\mathcal{S}}^{\mathsf{T}} \left(\mathbf{A}_{\mathcal{S}}^{\mathsf{T}}\mathbf{A}_{\mathcal{S}} - \mathbf{I}_s\right) \boldsymbol{x}_{\mathcal{S}}}{\|\boldsymbol{x}_{\mathcal{S}}\|_2^2} \leq \frac{\sigma_{\max}\left(\mathbf{A}_{\mathcal{S}}^{\mathsf{T}}\mathbf{A}_{\mathcal{S}} - \mathbf{I}_s\right) \|\boldsymbol{x}_{\mathcal{S}}\|_2^2}{\|\boldsymbol{x}_{\mathcal{S}}\|_2^2}$$

$$\leq \sigma_{\max}\left(\mathbf{A}_{\mathcal{S}}^{\mathsf{T}}\mathbf{A}_{\mathcal{S}} - \mathbf{I}_s\right)$$

As a result for any given $\mathcal{S}$, *we could say*

$$\max_{\boldsymbol{x}_{\mathcal{S}} \neq \mathbf{0}} \frac{\boldsymbol{x}_{\mathcal{S}}^{\mathsf{T}} \left(\mathbf{A}_{\mathcal{S}}^{\mathsf{T}}\mathbf{A}_{\mathcal{S}} - \mathbf{I}_s\right) \boldsymbol{x}_{\mathcal{S}}}{\|\boldsymbol{x}_{\mathcal{S}}\|_2^2} = \sigma_{\max}\left(\mathbf{A}_{\mathcal{S}}^{\mathsf{T}}\mathbf{A}_{\mathcal{S}} - \mathbf{I}_s\right)$$

Basic Definitions

How can we calculate the restricted isometry constant?

This concludes that

$$\delta_s(\mathbf{A}) = \max_{|\mathcal{S}|=s} \max_{\boldsymbol{x}_{\mathcal{S}} \neq \mathbf{0}} \frac{\boldsymbol{x}_{\mathcal{S}}^{\mathsf{T}} \left(\mathbf{A}_{\mathcal{S}}^{\mathsf{T}} \mathbf{A}_{\mathcal{S}} - \mathbf{I}_s \right) \boldsymbol{x}_{\mathcal{S}}}{\|\boldsymbol{x}_{\mathcal{S}}\|_2^2}$$

$$= \max_{|\mathcal{S}|=s} \sigma_{\max} \left(\mathbf{A}_{\mathcal{S}}^{\mathsf{T}} \mathbf{A}_{\mathcal{S}} - \mathbf{I}_s \right)$$

So, we could conclude that

$$\delta_s(\mathbf{A}) = \max_{|\mathcal{S}|=s} \sigma_{\max} \left(\mathbf{A}_{\mathcal{S}}^{\mathsf{T}} \mathbf{A}_{\mathcal{S}} - \mathbf{I}_s \right)$$

Basic Definitions

How complex is the calculation of restricted isometry constant?

Well, we need to solve

$$\delta_s(\mathbf{A}) = \max_{|\mathcal{S}|=s} \sigma_{\max}\left(\mathbf{A}_{\mathcal{S}}^{\mathsf{T}}\mathbf{A}_{\mathcal{S}} - \mathbf{I}_s\right)$$

This means that we should determine

$$\binom{N}{s}$$

cases and find maximum

Basic Definitions

How complex is the calculation of restricted isometry constant?

> *This means that similar to ℓ_1 coherence function*
>
> *Calculation of $\delta_s(\mathbf{A})$ is exponentially complex*

How we could check a constraint on restricted isometry constant?

> *We could always bound $\delta_s(\mathbf{A})$*

But before that, let's check the desired value for $\delta_s(\mathbf{A})$

Restricted Isometry Property

Section 1: Small Restricted Isometry Constant

Invertibility of Sampling

Assume we want to make sure that

samples collected by $\mathbf{A}$ *are invertible for any* s*-sparse signal*

From our discussions on coherence, we know that we need

$$\mu_1 (s-1) < 1$$

since eigenvalues of $\mathbf{A}_{\mathcal{S}}^{\mathsf{T}} \mathbf{A}_{\mathcal{S}}$ *for any* $\mathcal{S}$ *read*

$$1 - \mu_1 (s-1) \leq \lambda \leq 1 + \mu_1 (s-1)$$

Invertibility of Sampling

Assume we want to make sure that

samples collected by $\mathbf{A}$ *are invertible for any* s*-sparse signal*

So we could say that for any $\mathcal{S} = \mathrm{Supp}\left(\boldsymbol{x}\right)$, *we have*

$$\begin{aligned}\|\mathbf{A}\boldsymbol{x}\|_2^2 = \|\mathbf{A}_{\mathcal{S}}\boldsymbol{x}_{\mathcal{S}}\|_2^2 &\leq \lambda_{\max}\left(\mathbf{A}_{\mathcal{S}}^{\mathsf{T}}\mathbf{A}_{\mathcal{S}}\right)\|\boldsymbol{x}_{\mathcal{S}}\|_2^2 \\ &\leq \left[1 + \mu_1\left(s-1\right)\right]\|\boldsymbol{x}\|_2^2 \\ &< 2\|\boldsymbol{x}\|_2^2\end{aligned}$$

Invertibility of Sampling

Assume we want to make sure that

samples collected by $\mathbf{A}$ *are invertible for any s-sparse signal*

So we need any s-sparse signal $\boldsymbol{x}$ *to satisfy*

$$0\|\boldsymbol{x}\|_2^2 = 0 \leq \|\mathbf{A}\boldsymbol{x}\|_2^2 < 2\|\boldsymbol{x}\|_2^2$$

Now, let us rewrite it as

$$(1-1)\,\|\boldsymbol{x}\|_2^2 \leq \|\mathbf{A}\boldsymbol{x}\|_2^2 < (1+1)\,\|\boldsymbol{x}\|_2^2$$

Invertibility of Sampling

To make sure that

samples collected by $\mathbf{A}$ *are invertible for any* s*-sparse signal*

For any s*-sparse signal* $\boldsymbol{x}$

$$(1-1)\|\boldsymbol{x}\|_2^2 \leq \|\mathbf{A}\boldsymbol{x}\|_2^2 < (1+1)\|\boldsymbol{x}\|_2^2$$

So, we have

$$\delta_s(\mathbf{A}) \leq 1$$

Invertibility of Sampling

What is the conclusion?

> *The necessary and sufficient condition to make sure that samples collected by* $\mathbf{A}$ *are invertible for any* s*-sparse signal is the inequality*
>
> $$\delta_s(\mathbf{A}) < 1$$

We could in fact derive a more general result

Invertibility of Sampling

Bounding Spectrum of Matrices

For any support $\mathcal{S} \subset \{1, \dots, M\}$ with $|\mathcal{S}| = s$, the singular values of $\mathbf{A}_{\mathcal{S}}$, i.e., eigenvalues of $\mathbf{A}_{\mathcal{S}}^{\mathsf{T}} \mathbf{A}_{\mathcal{S}}$, read

$$1 - \delta_s(\mathbf{A}) \leq \lambda \leq 1 + \delta_s(\mathbf{A})$$

So, we need $\delta_s(\mathbf{A}) < 1$ to make sure that

by sampling via $\mathbf{A}$ we do not lose information on s-sparse signal

We look for matrices with small restricted isometry constant

Restricted Isometry Property

Restricted Isometry Property

A *satisfies the restricted isometry property or uniform uncertainty principle, if* $\delta_s(\mathbf{A})$ *remains relatively small, even if* s *grows large*

Note that this definition is not very formal!

In fact, restricted isometry property is not formally defined

Soon, we understand this property better!

Restricted Isometry Property

Section 2: Bounding Restricted Isometry Constant

Initial Observation

When we have deviation $\delta_s(\mathbf{A})$ for s-sparse signals

It guarantee that deviation $\leq \delta_s(\mathbf{A})$ for $(s-1)$-sparse signals

So, we could say in general

$$\delta_1(\mathbf{A}) \leq \delta_2(\mathbf{A}) \leq \ldots \leq \delta_N(\mathbf{A})$$

This means that $\delta_s(\mathbf{A})$ is increasing in s

A Trivial Bound

We have seen that for any $\mathcal{S}$

> $\delta_s(\mathbf{A})$ *is the minimum value for which*
>
> $$1 - \delta_s(\mathbf{A}) \leq \lambda\left(\mathbf{A}_{\mathcal{S}}^{\mathsf{T}}\mathbf{A}_{\mathcal{S}}\right) \leq 1 + \delta_s(\mathbf{A})$$

We also have seen that for any $\mathcal{S}$

> *For ℓ_1 coherence function, we have*
>
> $$1 - \mu_1(s-1) \leq \lambda\left(\mathbf{A}_{\mathcal{S}}^{\mathsf{T}}\mathbf{A}_{\mathcal{S}}\right) \leq 1 + \mu_1(s-1)$$

A Trivial Bound

Comparing the two results, it is clear that

$$\delta_s(\mathbf{A}) \leq \mu_1(s-1)$$

This bound is actually tight for the two first choices of s

$$\delta_1(\mathbf{A}) = 0 \qquad \textit{and} \qquad \delta_2(\mathbf{A}) = \mu_1(1) = \mu(\mathbf{A})$$

Noting that $\mu_1(s) \leq s\mu(\mathbf{A})$, *we could say*

$$\delta_s(\mathbf{A}) \leq (s-1)\,\mu(\mathbf{A})$$

A Non-Trivial Bound

To derive a lower-bound, we start with the following result

Ristricted Isometry Constant vs. Dimensions

Let $\mathbf{A} \in \mathbb{R}^{M \times N}$ *be a matrix with* $N \gg M$ *whose restricted isometry constant for sparsity* s *is* $\delta_s(\mathbf{A})$*. Then, its dimensions satisfy*

$$M \geq \frac{\frac{N}{s}(s-2)\left(1-\delta_s(\mathbf{A})\right)^2}{\frac{2N}{s-2}\delta_s^2(\mathbf{A}) + \left(1+\delta_s(\mathbf{A})\right)^2}$$

A Non-Trivial Bound

This results bounds

Minimum number of samples

in order to guarantee that

restricted isometry constant for sparsity s is $\delta_s(\mathbf{A})$

For $\delta_s(\mathbf{A})$, we need to have at least $M = M^\star$ samples

$$M^\star = \frac{\frac{N}{s}(s-2)(1-\delta_s(\mathbf{A}))^2}{\frac{2N}{s-2}\delta_s^2(\mathbf{A}) + (1+\delta_s(\mathbf{A}))^2}$$

A Non-Trivial Bound

Now assume that we use $\mathbf{A}$ *to sample a*

very sparse signal, i.e., $N \gg s$

in a high dimensional system, i.e.,

$$N, M, s \gg 1$$

To make sure that we can invert the samples, we need to set

$$0 \leq \delta_s(\mathbf{A}) < 1$$

A Non-Trivial Bound

Since $0 \leq \delta_s(\mathbf{A}) < 1$, *we could say*

$$0 \leq \delta_s^2(\mathbf{A}) < 1 \leq (1 + \delta_s(\mathbf{A}))^2 < 4$$

On the other hand, we could say

$$\frac{N}{s} \gg 1$$

So, we could conclude that

$$\frac{2N}{s-2}\delta_s^2(\mathbf{A}) + (1 + \delta_s(\mathbf{A}))^2 \approx \frac{2N}{s-2}\delta_s^2(\mathbf{A})$$

A Non-Trivial Bound

We further know that $s \gg 1$. *So, we could say*

$$s - 2 \approx s$$

This means that

$$\frac{N}{s}(s-2) \approx N$$

and we have

$$\frac{2N}{s-2}\delta_s^2(\mathbf{A}) + (1 + \delta_s(\mathbf{A}))^2 \approx \frac{2N}{s-2}\delta_s^2(\mathbf{A}) \approx \frac{2N}{s}\delta_s^2(\mathbf{A})$$

A Non-Trivial Bound

Now let us define the constant C to be

$$C = \frac{\left(1 - \delta_s\left(\mathbf{A}\right)\right)^2}{2}$$

Then the minimum required samples is approximately given by

$$\boxed{M^\star = \frac{\frac{N}{s}\left(s-2\right)\left(1 - \delta_s\left(\mathbf{A}\right)\right)^2}{\frac{2N}{s-2}\delta_s^2\left(\mathbf{A}\right) + \left(1 + \delta_s\left(\mathbf{A}\right)\right)^2} \approx C\frac{s}{\delta_s^2\left(\mathbf{A}\right)}}$$

A Non-Trivial Bound

So, we could conclude that

To make sure that we can do the recovery, we need

$$M \geq C \frac{s}{\delta_s^2 (\mathbf{A})}$$

samples!

This is in fact a general result!

A Non-Trivial Bound

Dimensions vs. Restricted Isometry Property

For $\mathbf{A} \in \mathbb{R}^{M \times N}$, let $N \geq UM$ and $\delta_s(\mathbf{A}) \leq \delta^\star$. Then, we can find a bounded constant C in terms of constants U and $\delta^\star$, such that minimum value of M varies in terms of $\delta_s(\mathbf{A})$ as

$$M \geq C \frac{s}{\delta_s^2(\mathbf{A})}$$

Now, let's see what do we get out of this result!

A Non-Trivial Bound

By simple modification, we get from this result that

For any matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, there is a bounded constant C

$$\delta_s(\mathbf{A}) \geq \sqrt{C \frac{s}{M}}$$

Combining with our trivial bound, we conclude that

$$\sqrt{C \frac{s}{M}} \leq \delta_s(\mathbf{A}) \leq (s-1)\,\mu(\mathbf{A})$$

A Non-Trivial Bound

Now assume we have such a constraint for successful recovery

$$\delta_s(\mathbf{A}) \leq \delta_0$$

where δ_0 is a constant less than 1, *e.g.*, $\delta_0 = 0.5$

What is the sufficient condition for recovery?

It is sufficient to have

$$(s-1)\,\mu(\mathbf{A}) \leq \delta_0$$

A Non-Trivial Bound

Now assume we have such a constraint for successful recovery

$$\delta_s(\mathbf{A}) \leq \delta_0$$

where δ_0 *is a constant less than* 1, *e.g.,* $\delta_0 = 0.5$

What is the sufficient condition for recovery?

In the best case, it reduces to a bound which says

$$M \geq C_0 s^2$$

which is as bad as the coherence bound!

A Non-Trivial Bound

Now assume we have such a constraint for successful recovery

$$\delta_s(\mathbf{A}) \leq \delta_0$$

where δ_0 is a constant less than 1, e.g., $\delta_0 = 0.5$

What is the necessary condition for recovery?

It is necessary to have

$$\sqrt{C\frac{s}{M}} \leq \delta_0$$

A Non-Trivial Bound

Now assume we have such a constraint for successful recovery

$$\delta_s(\mathbf{A}) \leq \delta_0$$

where δ_0 *is a constant less than* 1*, e.g.,* $\delta_0 = 0.5$

What is the necessary condition for recovery?

So this means that we need

$$M \geq C\delta_0^2 s$$

This is as good as optimal bound!

A Non-Trivial Bound

Now assume we have such a constraint for successful recovery

$$\delta_s(\mathbf{A}) \leq \delta_0$$

where δ_0 *is a constant less than* 1, *e.g.*, $\delta_0 = 0.5$

What do we conclude?

A bound on the restricted isometry property could lead to

much tighter sufficiency order

Restricted Isometry Property

Section 3: Recovery Guarantees for Basis Pursuit

Recovery Guarantee: Basis Pursuit

Remember this result ...

Recovery Guarantee by Null Space Property

Any s-sparse signal is uniquely recovered via basis pursuit algorithm from its samples collected by sampling matrix $\mathbf{A}$, if and only if $\mathbf{A}$ satisfies null space property of order s

This means that for any support $|\mathcal{S}| = s$, we need

$$\|\boldsymbol{v}_{\mathcal{S}}\|_1 < \|\boldsymbol{v}_{\bar{\mathcal{S}}}\|_1$$

for all $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$

Recovery Guarantee: Basis Pursuit

We can check the null space property differently

Let add $\|\mathbf{v}_{\mathcal{S}}\|_1$ *to both sides*

$$\underbrace{\|\mathbf{v}_{\mathcal{S}}\|_1 + \|\mathbf{v}_{\mathcal{S}}\|_1}_{2\|\mathbf{v}_{\mathcal{S}}\|_1} < \underbrace{\|\mathbf{v}_{\mathcal{S}}\|_1 + \|\mathbf{v}_{\bar{\mathcal{S}}}\|_1}_{\|\mathbf{v}\|_1}$$

So, we could say for all $\mathbf{0} \neq \mathbf{v} \in \ker \mathbf{A}$, *we need*

$$\|\mathbf{v}_{\mathcal{S}}\|_1 < \frac{1}{2}\|\mathbf{v}\|_1$$

and we should show it for any support $|\mathcal{S}| = s$

Recovery Guarantee: Basis Pursuit

We could further get rid of "any" support $|\mathcal{S}| = s$

Let us sort entries of $\boldsymbol{v}$ *as*

$$|v_{i_1}| > |v_{i_2}| > \ldots > |v_{i_N}|$$

and focus on support $\mathcal{S}_0 = \{i_1, \ldots, i_s\}$. *Clearly*

$$\|\boldsymbol{v}_{\mathcal{S}}\|_1 = \sum_{i \in \mathcal{S}} |v_i| \leq \sum_{i \in \mathcal{S}_0} |v_i| = \|\boldsymbol{v}_{\mathcal{S}_0}\|_1$$

for any support $|\mathcal{S}| = s$

Recovery Guarantee: Basis Pursuit

We could further get rid of "any" support $|\mathcal{S}| = s$

So, we could conclude that it is enough to show that

$$\|\boldsymbol{v}_{\mathcal{S}_0}\|_1 < \frac{1}{2}\|\boldsymbol{v}\|_1$$

for all $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$

Attention!

This does not reduce the complexity of null space property

We should find $\mathcal{S}_0$ *for every* $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$*!*

Recovery Guarantee: Basis Pursuit

We now start with this alternative form

$\mathbf{A}$ *satisfies the null space property, if*

$$\|\boldsymbol{v}_{\mathcal{S}_0}\|_1 < \frac{1}{2}\|\boldsymbol{v}\|_1$$

for all $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$

and find a sufficient condition for that in terms of

restricted isometry constant

Recovery Guarantee: Basis Pursuit

Consider a $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$ *and let's do the following decomposition*

$$\boldsymbol{v} = \sum_{k=0}^{K} \boldsymbol{v}_k$$

where $K = \lfloor N/s \rfloor$ *and*

$$\boldsymbol{v}_k = \begin{cases} (k+1)\text{-th } s\text{-largest entries of } \boldsymbol{v} & \text{for } k = 0, \dots, K-1 \\ \text{remained entries of } \boldsymbol{v} & \text{for } k = K \end{cases}$$

Let's check out an example together!

Recovery Guarantee: Basis Pursuit

Consider a $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$ *and let's do the following decomposition*

$$\boldsymbol{v} = \sum_{k=0}^{K} \boldsymbol{v}_k$$

Say we have $s = 2$ *and*

$$\boldsymbol{v} = \begin{bmatrix} 3 \\ -4 \\ 1 \\ 5 \\ -2 \end{bmatrix} \rightsquigarrow \boldsymbol{v}_0 = \begin{bmatrix} 0 \\ -4 \\ 0 \\ 5 \\ 0 \end{bmatrix} : \textit{ The first } s = 2 \textit{ largest entries of } \boldsymbol{v}$$

Recovery Guarantee: Basis Pursuit

Consider a $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$ *and let's do the following decomposition*

$$\boldsymbol{v} = \sum_{k=0}^{K} \boldsymbol{v}_k$$

Say we have $s = 2$ *and*

$$\boldsymbol{v} = \begin{bmatrix} 3 \\ -4 \\ 1 \\ 5 \\ -2 \end{bmatrix} \rightsquigarrow \boldsymbol{v}_1 = \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \\ -2 \end{bmatrix} : \textit{ The second } s = 2 \textit{ largest entries of } \boldsymbol{v}$$

Recovery Guarantee: Basis Pursuit

Consider a $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$ *and let's do the following decomposition*

$$\boldsymbol{v} = \sum_{k=0}^{K} \boldsymbol{v}_k$$

Say we have $s = 2$ *and*

$$\boldsymbol{v} = \begin{bmatrix} 3 \\ -4 \\ 1 \\ 5 \\ -2 \end{bmatrix} \rightsquigarrow \boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} : \textit{ The remained entries of } \boldsymbol{v}$$

Recovery Guarantee: Basis Pursuit

Consider a $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$ *and let's do the following decomposition*

$$\boldsymbol{v} = \sum_{k=0}^{K} \boldsymbol{v}_k$$

And we could write

$$\boldsymbol{v} = \begin{bmatrix} 3 \\ -4 \\ 1 \\ 5 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ -4 \\ 0 \\ 5 \\ 0 \end{bmatrix} + \begin{bmatrix} 3 \\ 0 \\ 0 \\ 0 \\ -2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Recovery Guarantee: Basis Pursuit

Consider a $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$ *and let's do the following decomposition*

$$\boldsymbol{v} = \sum_{k=0}^{K} \boldsymbol{v}_k$$

Clearly, in this decomposition

$$\boldsymbol{v}_0 = \boldsymbol{v}_{\mathcal{S}_0}$$

This means that we should show that for all $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$

$$\|\boldsymbol{v}_0\|_1 < \frac{1}{2}\|\boldsymbol{v}\|_1$$

Recovery Guarantee: Basis Pursuit

We start with the fact that $\boldsymbol{v} \in \ker \mathbf{A}$

We could write

$$\mathbf{0} = \mathbf{A}\boldsymbol{v} = \mathbf{A}\sum_{k=0}^{K} \boldsymbol{v}_k$$

Thus, we have

$$\mathbf{A}\boldsymbol{v}_0 = -\mathbf{A}\sum_{k=1}^{K} \boldsymbol{v}_k$$

Keep this in mind!

Recovery Guarantee: Basis Pursuit

Also keep in mind the following result

Bounding Inner Products

Let $\boldsymbol{x} \in \mathbb{R}^N$ be t-sparse and $\boldsymbol{y} \in \mathbb{R}^N$ be s-sparse. For any matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, we have

$$\boldsymbol{x}^{\mathsf{T}} \mathbf{A}^{\mathsf{T}} \mathbf{A} \boldsymbol{y} \leq \delta_{t+s}(\mathbf{A}) \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2$$

You get to prove it in Assignments 5!

Recovery Guarantee: Basis Pursuit

Now, let's start:

Since $\boldsymbol{v}_0$ is s-sparse, it is also $2s$-sparse!

By definition of restricted isometry constant, we have

$$(1 - \delta_{2s}(\mathbf{A})) \|\boldsymbol{v}_0\|_2^2 \leq \|\mathbf{A}\boldsymbol{v}_0\|_2^2$$

This means that

$$\|\boldsymbol{v}_0\|_2^2 \leq \frac{1}{1 - \delta_{2s}(\mathbf{A})} \|\mathbf{A}\boldsymbol{v}_0\|_2^2$$

Recovery Guarantee: Basis Pursuit

We could further write

$$\|\mathbf{A}\boldsymbol{v}_0\|_2^2 = (\mathbf{A}\boldsymbol{v}_0)^{\mathsf{T}} (\mathbf{A}\boldsymbol{v}_0)$$

and replace the second $\mathbf{A}\boldsymbol{v}_0$ *as*

$$\|\mathbf{A}\boldsymbol{v}_0\|_2^2 = (\mathbf{A}\boldsymbol{v}_0)^{\mathsf{T}} \left(-\mathbf{A} \sum_{k=1}^{K} \boldsymbol{v}_k \right) = - \sum_{k=1}^{K} \boldsymbol{v}_0^{\mathsf{T}} \mathbf{A}^{\mathsf{T}} \mathbf{A} \boldsymbol{v}_k$$

Since $\boldsymbol{v}_0$ *and* $\boldsymbol{v}_k$ *are both* s*-sparse, we could say*

$$-\boldsymbol{v}_0^{\mathsf{T}} \mathbf{A}^{\mathsf{T}} \mathbf{A} \boldsymbol{v}_k \leq \delta_{2s}(\mathbf{A}) \|\boldsymbol{v}_0\|_2 \|\boldsymbol{v}_k\|_2$$

Recovery Guarantee: Basis Pursuit

Thus, we have

$$\|\mathbf{A}\boldsymbol{v}_0\|_2^2 \leq \delta_{2s}(\mathbf{A}) \|\boldsymbol{v}_0\|_2 \sum_{k=1}^{K} \|\boldsymbol{v}_k\|_2$$

Combining with the earlier result, we have

$$\|\boldsymbol{v}_0\|_2^2 \leq \frac{1}{1-\delta_{2s}(\mathbf{A})} \|\mathbf{A}\boldsymbol{v}_0\|_2^2 \leq \frac{\delta_{2s}(\mathbf{A})}{1-\delta_{2s}(\mathbf{A})} \|\boldsymbol{v}_0\|_2 \sum_{k=1}^{K} \|\boldsymbol{v}_k\|_2$$

We can hence simplify it to

$$\|\boldsymbol{v}_0\|_2 \leq \frac{\delta_{2s}(\mathbf{A})}{1-\delta_{2s}(\mathbf{A})} \sum_{k=1}^{K} \|\boldsymbol{v}_k\|_2$$

Recovery Guarantee: Basis Pursuit

Now, we need to convert everything to ℓ_1-norm

We start by remembering the notation

$$|v_{i_1}| \geq |v_{i_2}| \geq \ldots \geq |v_{i_N}|$$

This means that

- $\boldsymbol{v}_0$ *contains* $v_{i_1}, \ldots, v_{i_s}$
- $\boldsymbol{v}_1$ *contains* $v_{i_{s+1}}, \ldots, v_{i_{2s}}$
- $\boldsymbol{v}_k$ *contains* $v_{i_{ks+1}}, \ldots, v_{i_{ks+s}}$

Recovery Guarantee: Basis Pursuit

Now, we need to convert everything to ℓ_1-norm

We hence could say

$$\|\mathbf{v}_k\|_2 = \sqrt{\sum_{j=1}^{s} \left| v_{i_{ks+j}} \right|^2} \leq \sqrt{s} \left| v_{i_{ks+1}} \right| \leq \sqrt{s} \left| v_{i_{ks}} \right|$$

$$\leq \sqrt{s} \left(\frac{\sum_{j=1}^{s} \left| v_{i_{(k-1)s+j}} \right|}{s} \right) = \sqrt{s} \left(\frac{\|\mathbf{v}_{k-1}\|_1}{s} \right)$$

$$= \frac{\|\mathbf{v}_{k-1}\|_1}{\sqrt{s}}$$

Recovery Guarantee: Basis Pursuit

Now, we need to convert everything to ℓ_1-norm

This means that

$$\|\mathbf{v}_0\|_2 \leq \frac{\delta_{2s}(\mathbf{A})}{1-\delta_{2s}(\mathbf{A})} \sum_{k=1}^{K} \|\mathbf{v}_k\|_2 \leq \frac{\delta_{2s}(\mathbf{A})}{1-\delta_{2s}(\mathbf{A})} \sum_{k=1}^{K} \frac{\|\mathbf{v}_{k-1}\|_1}{\sqrt{s}}$$

$$\leq \frac{1}{\sqrt{s}} \left(\frac{\delta_{2s}(\mathbf{A})}{1-\delta_{2s}(\mathbf{A})} \right) \sum_{k=0}^{K-1} \|\mathbf{v}_k\|_1$$

$$\leq \frac{1}{\sqrt{s}} \left(\frac{\delta_{2s}(\mathbf{A})}{1-\delta_{2s}(\mathbf{A})} \right) \sum_{k=0}^{K} \|\mathbf{v}_k\|_1 = \left(\frac{\delta_{2s}(\mathbf{A})}{1-\delta_{2s}(\mathbf{A})} \right) \frac{\|\mathbf{v}\|_1}{\sqrt{s}}$$

Recovery Guarantee: Basis Pursuit

We further use the following result

> *For any s-sparse signal $\mathbf{x}$, we have*
>
> $$\|\mathbf{x}\|_2 \geq \frac{\|\mathbf{x}\|_1}{\sqrt{s}}$$

You will prove it via Jensen's inequality in Assignments 5!

> *Since $\mathbf{v}_0$ is s-sparse, we have*
>
> $$\frac{\|\mathbf{v}_0\|_1}{\sqrt{s}} \leq \|\mathbf{v}_0\|_2 \leq \left(\frac{\delta_{2s}(\mathbf{A})}{1 - \delta_{2s}(\mathbf{A})} \right) \frac{\|\mathbf{v}\|_1}{\sqrt{s}}$$

Recovery Guarantee: Basis Pursuit

What do we conclude?

For any $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$, we have

$$\|\boldsymbol{v}_0\|_1 \leq \frac{\delta_{2s}(\mathbf{A})}{1-\delta_{2s}(\mathbf{A})} \|\boldsymbol{v}\|_1$$

What do we need to have null space property?

For any $\mathbf{0} \neq \boldsymbol{v} \in \ker \mathbf{A}$, we have

$$\|\boldsymbol{v}_0\|_1 < \frac{1}{2} \|\boldsymbol{v}\|_1$$

Recovery Guarantee: Basis Pursuit

So, it is sufficient to have

$$\frac{\delta_{2s}(\mathbf{A})}{1-\delta_{2s}(\mathbf{A})} < \frac{1}{2}$$

Or equivalently

$$\delta_{2s}(\mathbf{A}) < \frac{1}{3}$$

Recovery Guarantee: Basis Pursuit

Recovery Guarantee for Basis Pursuit

Any s-sparse signal is recovered from samples collected by matrix $\mathbf{A}$ via the basis pursuit algorithm, if

$$\delta_{2s}(\mathbf{A}) < \frac{1}{3}$$

Restricted Isometry Property

Section 4: Other Recovery Guarantees

Recovery Guarantee: General Form

Other recovery guarantees look the same

A given algorithm is guaranteed to

- *either recover exactly in noise-free cases*
- *or estimate true signal with bounded error in noisy cases*

from the samples collected by $\mathbf{A}$ *if we have*

$$\delta_{\kappa s}(\mathbf{A}) < \delta^{\star}$$

for some integer κ *and constant* $\delta^{\star}$

Recovery Guarantee: OMP

For instance for OMP, we have

Recovery Guarantee for OMP

Any s-sparse signal is recovered from samples collected by matrix $\mathbf{A}$ via OMP algorithm, if

$$\delta_{13s}(\mathbf{A}) < \frac{1}{6}$$

Recovery Guarantee: IHT

For iterative hard thresholding also we have

Recovery Guarantee for IHT

Any s-sparse signal is recovered from samples collected by matrix $\mathbf{A}$ *via IHT algorithm, if*

$$\delta_{3s}(\mathbf{A}) < \frac{1}{2}$$

Results on the noisy cases can be further checked in Chapter 6

Recovery Guarantee: Sufficiency Order

Such recovery guarantees result in much smaller sufficiency order

Sufficiency Order of RIP

Recovery guarantees based on the restricted isometry property of a matrix determine the sufficient number of samples for recovery as

$$M \geq Cs$$

for some constant C

Recovery Guarantee: Sufficiency Order

But can we construct a matrix which satisfies RIP?

Well! Deterministically it is hard, but we can do it randomly

How is it done randomly?

Simply generate $\mathbf{A} \in \mathbb{R}^{M \times N}$ *at random with*

$$M \geq 2s \ln \left(\frac{N}{s} \right)$$

rows. Then, with high probability $\mathbf{A}$ *satisfies RIP*

Final Points

Summary

In a nutshell, we have learned

- *Null space property*
  - *Give necessary and sufficient conditions for recovery*
  - *It is computationally hard to be checked*
- *Alternatively, we can find guarantees in terms on coherence*
  - *We get rid of high complexity issue*
  - *The results give a loose sufficiency order*
- *Better guarantees are given via restricted isometry constant*
  - *Some sufficient bounds are derived tractably*
  - *The results show a tighter sufficiency order*

Which Parts of Textbooks?

We are over with this part

I would suggest to go over

A Mathematical Introduction to Compressive Sensing
S. Foucart and H. Rauhut, Book, 2013

and study the following parts:

- *Null Space Property: Chapter 4: Till end of Section 4.3*
- *Coherence: Chapter 5*
- *Restricted Isometry Property: Chapter 6*